8-9-2019

Investigation of Antimicrobial Activity and Metabolic Genes in Bacterial Symbionts Associated with the Female Reproductive System of the Hawaiian Bobtail Squid, Euprymna scolopes

Andrea Suria
University of Connecticut - Storrs, andrea.suria@uconn.edu

Investigation of Antimicrobial Activity and Metabolic Genes in Bacterial Symbionts Associated with the Female Reproductive System of the Hawaiian Bobtail Squid, *Euprymna scolopes*

Andrea Maria Suria, Ph.D.

University of Connecticut, 2019

The Hawaiian bobtail squid, *Euprymna scolopes*, maintains a diverse bacterial consortium in epithelium lined tubules in a female reproductive organ, the accessory nidamental gland (ANG). During egg laying, bacteria from the ANG are deposited into a jelly coat (JC) that surrounds the yolk sac. The ANG/JC communities are composed of *Alphaproteobacteria*, *Verrucomicrobia*, *Flavobacteriia*, and *Gammaproteobacteria*. Removing or altering the JC bacterial community leaves eggs vulnerable to biofouling and infections. To understand the metabolism of this community and look for potential mechanisms of egg defense, the antimicrobial activity of ANG/JC symbionts was tested, 4 ANG metagenomes were sequenced, and metatranscriptomes were sequenced from 6 ANGs, 6 freshly laid egg JCs, 5 JCs from eggs midway through development, and 5 JCs that were challenged with a fungal pathogen for half of egg development. Inhibition against Gram-negative and/or Gram-positive marine bacteria was seen in 20% of tested isolates, and strong to moderate activity against the fungal pathogen, *Fusarium keratoplasticum*, was seen in 33% of tested isolates *in vitro.* Metagenomic sequencing of the ANG resulted in recovery of 9 new high quality genomes from *Verrucomicrobia, Alterythrobacter, Mesorhizobium,* and *Vibrio* strains. Mapping metatranscriptomic reads to one *Verrucomicrobia* genome revealed utilization of many nutrients in the ANG and JCs, including

fucose, sulfate, and sialic acid, likely originating from host-produced glycoproteins. Analysis of the ANG metagenome also predicted 255 secondary metabolite biosynthetic gene clusters, which may produce or regulate some antimicrobial compounds that could act in egg defense. Of these clusters, genes for synthesis of nonribosomal peptide synthetases (NRPSs), polyketide synthases (PKSs), terpenes, bacteriocins, thiopeptides, homoserine lactones, siderophores, and aryl polyenes were expressed at similar levels between all ANG and JC samples. The antimicrobial compounds, indigoidine and bromoalterochromides, were described in two JC isolates, *Leisingera* sp. JC1 and *Pseudoalteromonas* sp. JC28. Expression of secondary metabolite genes in the ANG and eggs demonstrated the potential for other defensive bioactive compounds to be produced and provides a baseline to further explore the mechanisms of bacterial egg defense in *E. scolopes*.

Investigation of Antimicrobial Activity and Metabolic Genes in Bacterial Symbionts Associated with the Female Reproductive System of the Hawaiian Bobtail Squid, *Euprymna scolopes*

Andrea Maria Suria

B.S., Fairleigh Dickinson University, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

At the

University of Connecticut

2019

APPROVAL PAGE

Doctor of Philosophy Dissertation

Investigation of Antimicrobial Activity and Metabolic Genes in Bacterial Symbionts Associated with the Female Reproductive System of the Hawaiian Bobtail Squid, *Euprymna scolopes*

Presented by

Andrea Maria Suria, B.S.

Major Advisor _____
Spencer Nyholm

Associate Advisor _____
Joerg Graf

Associate Advisor _____
Marcy Balunas

Associate Advisor _____
Jonathan Klassen

Associate Advisor _____
Nichole Broderick

University of Connecticut
2019

Acknowledgements

I would first like to thank my advisor, Dr. Spencer Nyholm, who gave me the freedom to try and fail, the support to keep trying, and the opportunity to learn so many new things, travel to many new places, and meet many amazing people. I thank my committee members, Dr. Marcy Balunas, Dr. Nichole Broderick, Dr. Daniel Gage, Dr. Joerg Graf, and Dr. Jonathan Klassen, whose valuable suggestions have made this research possible and challenged me to think broadly and critically. My collaboration with the Balunas lab especially allowed for many exciting interdisciplinary findings between biology and chemistry.

I thank the current and previous Nyholm lab members, who fostered a supportive environment to share ideas and adventures. I would especially like to thank Dr. Allison Kerwin and Dr. Andrew Collins, who have taught me so much about molecular biology and the ANG symbiosis. Their continued support at every stage of my Ph.D. has helped me push forward and overcome many challenges. I thank Dr. Artemis (Dyanna) Louyakis for guiding me through my metagenomic/metatranscriptomic projects and always helping me with bioinformatic errors. I thank my academic siblings, Sarah McAnulty and Michael Stephens, who have been by my side every step of this journey. We have shared all the highs and lows possible these last six years and I could not have asked for more enthusiastic friends to make wild dreams possible. I thank the many undergrads who helped take care of the squid that made this research possible. And to the undergrads I personally mentored—Hebo Ismail, Jessica Bertenshaw, Jaydeen Sewell, Lydia Abini-Agbomson, Sarah Cleveland, and Abishek Arokiadoss—thank you for allowing me to be a part of your journeys. With each one of you I have learned to be a more patient and understanding mentor.

To my family, you have provided me continuous support and unwavering belief in this dream. All of my strength and resilience, I owe to you. I know that each step I take forward, I take for all of us. *Sin todos ustedes, no podría estar aquí. Con todo mi corazón, muchismas gracias.*

When I was five years old, I told my parents that I wanted to be a marine biologist, go to Hawaii, and discover three new fishes. With this thesis, I can say that I grew up to be a marine microbiologist, I went to Hawaii, and isolated many new bacteria. And it has been a more challenging and fulfilling experience than I could have ever imagined.

Table of Contents

List of Tables

Chapter Four

Appendix I

List of Figures

Chapter Two – Part One

Chapter Two – Part Two

Chapter Three

Chapter Four

Appendix I

Appendix II

Chapter 1

Introduction

All organisms, from single-celled protists to multi-cellular plants and animals, live in association with microbes (1). These symbiotic relationships exist on a spectrum of mutualism to parasitism. Symbioses that provide benefits to the host can generally be categorized as either nutritional, where microbes provide essential nutrients not found in the host's diet (2), or defensive, where microbes provide either direct or indirect host protection from predators, parasites, or pathogens. Nutritional symbioses are often obligate associations, where symbionts are vertically transmitted from parents to offspring. In some cases this has resulted in genome co-evolution, such as the reduction of symbiont genomes to rely on host genomes for completion of metabolic pathways (3). It should be noted that there are exceptions to this mode of transmission, such as the obligate, horizontally-acquired sulfur-oxidizing endosymbionts of deep sea tubeworms (4). Defensive symbioses can be more flexible, where phylum-level taxa may be conserved with greater diversity in lower taxonomic levels (5). These associations are more frequently horizontally transmitted, where each generation must acquire their symbionts from the environment. Symbiont strain metabolism can vary between host species and host populations, potentially to adapt to local threats (6).

The host protection provided by defensive symbionts can occur through competitive exclusion, indirect priming of the host's immune system, or direct production of bioactive compounds. Competitive exclusion, or colonization resistance, in the mammalian gut occurs when a stable, dense microbial community can prevent invasion of potential pathogens (7). When this microbial community is disturbed, such as by antibiotics, the opportunity for a pathogen to invade or dominate the community increases. For example, in humans this can result in *Clostridium difficile* infections, the causative agent of pseudomembranous colitis, but treatment with fecal microbiota transplantation from a healthy donor can restore the normal gut

community within the next day (8). The colonization of a host by microbes can be necessary to prime certain aspects of the host's immune system. Colonization of the roots of many plants by bacteria and arbuscular mycorrhizal fungi can induce the plant to produce more cellular defenses in response to pathogen attack in distant parts of the plant, such as in the leaves (9). In the mosquito, *Aedes aegypti*, infection with *Wolbachia w*AlbB induces resistance to the dengue virus through up-regulation of the host's NADPH oxidase and production of reactive oxygen species, which in turn increases the host's production of antimicrobial peptides, cecropins and defensins (10). Studies from axenic mice have shown that a conventional gut microbiome is necessary for proper development of Peyer's patches in the small intestine, mesentery lymph nodes, and regular levels of T cells, IgA, and other immune factors (11).

There have been a number of organisms across the tree of life where the origin of defensive compounds can be assigned to symbiotic microbes. The sponge, *Theonella swinhoei*, is protected by the cytotoxic macrolide, swinholide A, produced by unidentified heterotrophic bacteria (12)(13), and an antifungal glycopeptide, theopalauamide, produced by '*Candidatus* Entotheonella palauensis' (14)(15). A bacterium isolated from salamander skin, *Janthinobacterium lividum*, produces indole-3-carboxaldehyde and violacein, which can inhibit the fungal skin pathogen, *Batrachochytrium dendrobatidis* (16). Fungus-farming ants possess *Pseudonocardia* and *Streptomyces* bacteria on their cuticle, which produce antifungals such as candicidin, dentigerumycin, and a nystatin-like polyene, which inhibit a fungal pathogen of the fungus gardens that the ants maintain as a food source (17)(18). While there are many examples of microbes protecting the adult life stages of organisms, there have been fewer examples of microbes chemically defending eggs or larval stages, especially in aquatic environments.

Defensive Symbioses in Eggs/Larva

Production and survival of offspring must occur in the face of many biological and environmental stressors, such as predation, climate change, or risk of infection (19)(20). Animals that deposit eggs in soil must contend with 10^8 -10^9 bacteria and 10^5 -10^6 fungi per gram of soil (21), and eggs in marine environments face 10^6 bacteria per milliliter of seawater (22). Particularly for aquatic eggs, biofilms can form on biotic or abiotic surfaces within minutes (23), and eggs in gelatinous masses develop at the limit of oxygen availability, so any additional biomass could further deplete oxygen from the embryos (24). Many animals display protective behaviors to defend their eggs, such as guarding/carrying, cleaning, or aerating eggs (25)(26). Animals that do not provide parental care may use a combination of egg structure (27)(28), provision of antimicrobial peptides (29)(30), or symbiotic microbes to defend their eggs (31).

Terrestrial egg/larva defense

Studies of chemical ecology and symbiosis in terrestrial systems have revealed several examples of organisms that use the chemicals derived from their symbionts to protect their offspring. Entomopathogenic nematodes have a complex life cycle and benefit from microbial compounds at all stages. Nematodes of the *Steinernematidae* family are colonized by *Xenorhabdus* sp. and *Heterorhabditidae* nematodes are colonized by *Photorhabdus* sp. (32). Both groups of nematodes enter an insect host during their infective juvenile stage, kill the host, and then mature and reproduce in the same host until all resources have been utilized, at which point infective juveniles emerge to find a new host. While several secondary metabolites produced by the symbionts contribute to killing the insect, many other compounds are involved in preserving the cadaver for the maturing nematodes. Both *Xenorhabdus* and *Photorhabdus*

produce rhabduscin and chitinases, which aid in insect killing, but also deter fungal competitors (33)(34)(35). Some small molecules that defend the cadaver from bacterial competitors include bacteriocins, stilbene, nematophin, indole derivatives, and anthraquinone derivatives (36). An ant-deterrent factor also prevents ants from feeding on the insect cadaver (37).

Some beetle species utilize microbes to defend their eggs from infection and predation. Female rove beetles, *Paederus* spp., have a facultative symbiosis with a *Pseudomonas* sp. which can produce pederin (38). Symbiotic females pass their pederin-producing symbionts into their eggs, and larvae that possess pederin deter spider predators (39). The *Lagria villosa* beetles possess different *Burkholderia gladioli* strains that are vertically transmitted in their eggs. These *B. gladioli* strains make several compounds that protect the eggs from fungal infections, including lagriamide (40), toxoflavin, caryoynensin, lagriene, and sinapigladioside (41). The *Euops* leaf-rolling weevils deposit their eggs and spores of an ectosymbiont, *Penicillium herquei*, in leaf cradles that develop on the soil floor. The *P. herquei* fungus makes (+)-scleroderolide, which inhibits growth of filamentous fungi, yeast, and bacteria normally found on leaves used to create cradles (42).

Other insects with known microbial chemical defense of their offspring include bees, wasps, and flies. Eusocial honey bees, *Apis mellifera,* feed their larvae bee bread, a mixture of pollen and nectar fermented by lactic acid bacteria, *Lactobacillus* and *Bifidobacterium*, in the worker bee's honey stomach. Although the larvae are not colonized until they emerge as adults (43), bacteria in their diet may protect them from infections. *Paenibacillus larvae* spores cause lethal infections in larvae, but one study has shown that significantly fewer larvae die when fed a mixture of *P. larvae* and lactic acid bacteria, compared to larvae fed *P. larvae* alone (44). The exact mechanism behind this defense is unknown but could be related to the antimicrobials lactic

acid bacteria produce, such as organic acids, hydrogen peroxide, and antimicrobial peptides. A predator of bees, solitary beewolf wasps, *Philanthus* spp., possess specialized antennal glands that house the actinobacteria, '*Candidatus* Streptomyces philanthi' (45). Females dig underground burrows to lay their eggs and apply secretions from these glands to their brood cells, which the larvae also incorporate into their cocoons. This symbiont produces streptochlorin and a complex of eight piericidin derivatives (46) that prevent fungal and microbial infections during the extended period of larval development, which can last up to nine months (47)(48). Eggs of the housefly, *Musca domestica*, are colonized by a mixture of 19 bacterial strains (*Gammaproteobacteria, Betaproteobacteria, Bacilli, Flavobacteriia*) which collectively inhibit fungal strains isolated from chicken manure where the eggs are laid (49). It is still uncertain, however, if this defense is a result of colonization resistance or production of secondary metabolites.

One of the few described examples in a vertebrate, hoopoe birds, *Upupa epops*, apply preen gland oil containing symbiotic bacteria to their eggs for protection (50). The uropygial oil of nesting hoopoes is dark-brown and malodorous, owing to a mixture of volatile compounds produced by bacteria in the oil that are absent in oil from birds treated with antibiotics (51). There are ten identified compounds in this oil, including butanoic acid, 2-methyl butanoic acid, benzaldehyde, phenol, 4-methyl pentanoic acid, phenyl acetaldehyde, indole, 3-phenyl propanoic acid, and 4-chloro indole—many of which can inhibit several Gram negative and Gram positive bacteria *in vitro* (51). All isolated bacteria from the uropygial gland have been *Enterococcus* spp., predominantly *E. faecalis*. Antimicrobial activity amongst the strains seems to be correlated with the presence of bacteriocin genes, particularly enteroicins MR10 and AS-48 (52).

It has also been shown that hoopoe eggshells have specialized crypts which become filled with the enterococci bacteria from the preen oil. When access to the gland was impeded, overall bacterial load on the eggshells and hatching success decreased (53). The strong odor produced by the bacterial volatiles of the preen oil may also prevent predation.

Marine egg/larva defense

There are far fewer examples of symbiont-produced chemical defense amongst marine organisms. The earliest published example is in the shrimp, *Palaemon macrodactylus*, which brood eggs densely covered by the bacterium, *Alteromonas* sp. I-2. This bacterium produces 2,3-indolinedione (isatin), which inhibits the known crustacean fungal pathogen, *Lagenidium callinectes* (54). Similarly, the eggs of the lobster, *Homarus americanus*, are covered by a salmon-colored, rod-shaped, Gram negative bacterium that produce 4-hydroxyphenethyl alcohol (tyrosol), which also inhibits *L. callinectes* (55). Larvae of the bryozoans, *Bugula neritina* and *Bugula simplex*, contain the bacteria '*Candidatus* Endobugula sertula' (56) and '*Candidatus* Endobugula glebosa' (57), respectively, in a circular invagination on their surface called the pallial sinus. These bacteria have been associated with production of bryostatins (58)(57), a family of 20 known bioactive compounds with anticancer activity (59). Extracts from larval *B. neritina* contain a mix of bryostatins which deter pinfish predation (60). In larvae of the Caribbean tunicate, *Trididemnum solidum*, there is correlative evidence of chemical defense by a symbiont. Blue-head wrasse consume significantly less *T. solidum* larvae when compared to a larval mimic (krill eyes), and extracts of didemnin B and nordidemnin B isolated from these larvae also deter predation (61). The source of these didemnin compounds is thought to be symbiotic cyanobacteria (62), which can produce compounds with similar structures, but no

study has directly linked tunicate-isolated cyanobacteria with didemnin production. Didemnin B, however, is produced by free-living *Alphaproteobacteria*, *Tistrella mobilis* and *Tistrella bauzanensis* (63), so there is evidence of a bacterial origin for this compound. Although there are other tunicates which have had the source of bioactive compounds associated with symbiotic bacteria (reviewed in (64)), there are no discussions of the roles of these compounds in larval defense.

The Accessory Nidamental Gland (ANG) Symbiosis

One symbiosis which has only recently been assigned a defensive function is between cephalopods and a bacterial consortium in the accessory nidamental gland (ANG, Appendix I). The ANG is part of the female reproductive system of some squid, bobtail squid, and cuttlefish (orders: *Sepiida*, *Sepiolida*, *Myopsida*) (65). The ANG is comprised of many intertwined, epithelium-lined tubules that are densely packed with bacteria. The openings of these tubules collect in close proximity to the openings of the nidamental glands, such that the contents from both glands combine to form the jelly coat (JC) that surrounds the embryo. The secretions from the nidamental gland compose the jelly-like matrix, while the ANG contributes bacteria embedded in the jelly (65). In one of the earliest descriptions of cephalopod ANGs, tubules with different pigmentation were described containing seemingly uniform cultures of different bacterial types, with bacilli in white and yellow tubules and cocci in red tubules in the European cuttlefish, *Sepia officinalis*, and bacilli in red tubules and cocci in white tubules in the bobtail squid, *Sepiola intermedia* (66). This observation was also reported in *Doryteuthis pealeii* (formerly *Loligo pealeii*) using electron microscopy (67). Bacteria of uniform shape (cocci or bacilli) were observed in different tubules, and were hypothesized to produce different pigments

that contribute to the coloration of individual tubules (67). Fluorescence in situ hybridization (FISH) was later used to confirm that different tubules are indeed colonized by different bacterial taxa in *Doryteuthis pealeii* (68) and *Euprymna scolopes* (69). Transmission electron micrographs (TEMs) in *E. scolopes* also showed that tubules containing different bacterial morphotypes have different epithelia (69). Tubules that contained small bacilli bacteria had vacuole-rich epithelia with electron-dense granules at the apical surface, while tubules with large coccoid bacteria had electron-dense epithelia lacking vacuoles. These differences suggest that the host may supply different nutrients in the different tubule microenvironments.

Early descriptions of bacteria in the eggs of cephalopods containing ANGs can also be accredited to the Italian zoologist, Umberto Pierantoni. Buchner describes that Pierantoni observed bacteria in the jelly layer of *Sepia* eggs which were similar to the bacteria found in the ANG (65). In recent studies, the first description of bacteria in cephalopod eggs, originating from the ANG, is in *Doryteuthis opalescens* (formerly *Loligo opalescens*) (70). TEMs of freshly laid *D. opalescens* eggs showed that the jelly layer is 1-2 mm thick and contains approximately 100 straited layers that are likely high in protein content. Bacteria were found in between these striated layers with an estimated 10^6 - 10^8 bacteria per egg (70). A study of *D. pealeii* ANGs and eggs using 16S rRNA gene clone libraries and FISH found that similar *Alphaproteobacteria* and *Gammaproteobacteria* were found throughout the jelly layer of egg capsules as in the ANG tubules (68). The same methods were used to identify similar bacteria in the eggs and ANGs of *Idiosepius biserialis*, *Doryteuthis vulgaris*, and *Sepioteuthis lessoniana* (71). FISH was used with *E. scolopes* eggs to show that the bacterial taxa found in the ANG are mixed amongst the jelly layers and do not show the same partitioning as in the ANG (69). Sequencing of the 16S rRNA gene in *E. scolopes* eggs and ANGs showed a high similarity in the operational taxonomic units

(OTUs) between these communities, and clustering of a female's ANG to her respective eggs (72).

The ANG/egg bacterial communities vary by cephalopod species but are all dominated by *Alphaproteobacteria*. Sequencing of 16S rRNA gene clone libraries have largely been used to identify these bacteria. The majority of clones from the cuttlefish, *Sepia officinalis*, had similarity to *Agrobacterium*, followed by *Roseobacter*, *Rhodium-Xanthobacter*, and the Gram-positive *Sporichthya* and *Clostridium* (73). All clones from the ANG and eggs of *D. pealeii* were identified as *Alphaproteobacteria* (SAR 83, *Roseobacter*, *Rhizobium*) and *Magnetobacterium*, while the majority of cultured isolates were *Gammaproteobacteria* (*Pseudoalteromonas*, *Shewanella*, *Vibrio*, *Colwellia*) (68). A survey of European and Western Pacific squid sequenced 128 clones from 10 squid (71). In this survey, *Photololigo* spp. ANGs were dominated by *Alphaproteobacteria* (72-80% of clones, mostly *Agrobacterium-Silicibacter*, *Rhodobacter*), *Cytophagia-Flavobacteriia-Bacteroidetes* (CFB, 0-11% of clones), and *Gammaproteobacteria* (5-26% of clones). *Loligo* spp. ANGs had *Alphaproteobacteria* (50-95% clones, mostly *Rhodobacter*), CFB (0-33% clones), *Gammaproteobacteria* (14-17% clones), and *Clostridium* (0-5% clones). *Loliolus* spp. ANGs have *Alphaproteobacteria* (49-88% clones, mostly *Rhodobacter*, *Roseobacter*), CFB (0-39% clones), and *Gammaproteobacteria* (12%). The ANGs of *Sepioteuthis lessoniana* had 100% *Alphaproteobacteria* (mostly *Rhodobacter*) and of *Idiosepius pygmaeus* had 92% *Alphaproteobacteria* (mostly *Rhodobacter*) and 8% *Clostridium*. In this same survey, *S. lessoniana* and *L. vulgaris* eggs had clones similar to *Agrobacterium*, *Roseobacter*, and *Rhodobium* species (71). The first metagenomic survey of an ANG was in *Eupyrmna scolopes*, which was composed of 72.6% *Alphaproteobacteria*, 16.7% *Verrucomicrobia*, 10.3% *Bacteroidetes*, and 0.37% *Gammaproteobacteria* (69). This is the first

report of *Verrucomicrobia* as a member of an ANG community. High-throughput sequencing of the 16S rRNA gene in *E. scolopes* confirmed that *Alphaproteobacteria* make up the greatest average abundance community member in both the ANG and eggs (66%/71.3% respectively), followed by the *Verrucomicrobia* (29.1%/11.5%), *Gammaproteobacteria* (3.8%/4.1%), and *Flavobacteriia* (0.9%/12.3%) (72).

Since the first observations of bacteria in ANGs and eggs, there have been several hypotheses for the function of these bacterial communities in cephalopods. Pierantoni first suggested that the bacteria might send out "rays" which have a positive influence on embryo development (65). It was also suggested that the bacteria may have some influence on a female's reproductive state, since ANG coloration changes from white to mottled red-orange with sexual maturity due to bacterial pigmentation (67)(74)(75). Researchers considered if bacteria were deposited into the eggs for vertical transmission, but due to the low density of bacteria in the eggs, this was considered unlikely (75). It should be noted that ANG tissue does not begin to develop until at least 87 days after hatching in *D. opalescens* squid (76), so there does not appear to be a dedicated tissue present in juveniles that would be immediately colonized. As well, no ANG/egg bacteria have been detected in *E. scolopes* hatchlings (77). Biggs and Epel (70) were the first to propose that the jelly coat bacteria may serve a defensive function for the eggs through either competitive exclusion or production of defensive chemicals. This hypothesis stemmed from divers' observations that *D. opalescens* eggs are rarely subject to predation or fouling by bacteria, fungi, or animals (70). *D. opalescens* eggs that were placed on hooks in Monterey, CA, were not preyed upon during the time that adult squid and fish on adjacent hooks were completely consumed (78). Eggs placed in aquaria were not consumed by starved sea stars, crabs, hermit crabs, isopods, moon snails, or large anemones (78). Video recordings of *L.*

vulgaris breeding off the south-eastern Cape of South Africa have shown pajama sharks removing a few egg strands from large egg masses, but eggs were not consumed (79). Despite many potential predators observed in this area, only the sparid fish, *Spondylisoma emarginatum,* were rarely observed feeding on egg strands (80). In contrast to squid and cuttlefish eggs, unattended octopus eggs are often consumed by predators and experience algal, fungal, or bacterial infections (71). Incirrate octopuses lack nidamental glands and accessory nidamental glands, and their eggs are composed of a yolk sac surrounded by a chorion, lacking any outer jelly layers (81). Due to this lack of egg encapsulation, female octopuses brood or carry their eggs to provide parental protection over the course of embryogenesis.

Nearly all studies that attempted to address the potential defensive function of ANG/egg bacteria utilized *in vitro* antimicrobial assays. Extracts from ANG tissue, egg JCs, or bacterial isolates were tested in zone of inhibition assays against Gram positive and Gram negative bacteria and one fungus, however there is no evidence that any of these microbes are potential cephalopod egg pathogens. Barbieri et al. (82) were the first to perform these tests, using ANG/JC/isolate butanol extracts from *D. pealeii*. It was found that ANG extracts could inhibit *Streptomyces griseus*, *Vibrio anguillarum*, and *Aeromonas salmonicida*, but egg extracts did not have antimicrobial activity. Of the isolate extracts, a *Shewanella* sp. inhibited *V. anguillarum*, *A. salmonicida*, *S. aureus,* and the fungus *Laegenidium myophilum*; an *Altermonas* sp. inhibited *E. coli, S. griseus, V. anguillarum, A. salmonicida,* and *L. myophilum*; and a *Roseobacter* sp. did not inhibit any of the target strains. Organic extracts from the cuttlefish *Sepia pharaonis*, *Sepia aculteata*, and the squid *Loligo duvaucelii* inhibited *E. coli*, *Aeromonas hydrophila*, *S. aureus*, and *B. megaterium*, while extracts from *Sepiella intermis* only inhibited *E. coli* and *B. megaterium* (83). Aqueous extracts from this same study did not have antibacterial activity, and

they noted that peak activity in organic extracts corresponded with the peak breeding period for each species. From this observation, it was later found that the difference in antibacterial activity of extracts from immature versus mature ANGs is due to a higher free fatty acid content in mature glands, particularly due to unsaturated fatty acids (84). *E. scolopes* ANG and egg isolates and their organic extracts have been shown to inhibit vibrio species and the fungal pathogen, *Fusarium keratoplasticum* (85) (see Chapter 1 and Appendix I). While these studies do not clearly divide antimicrobial activity associated with the host or symbiotic bacteria, there is one report that the jelly itself of *D. pealeii* eggs contains a cilia inhibitory factor which can cause asynchronous beat and potentially immobilize larva or protists (86).

A defensive function of the ANG symbiosis was determined *in vivo* in *E. scolopes* using antibiotic treatment of eggs (Appendix I). Freshly laid egg clutches were separated into an untreated, control group and an experimental group which received an antibiotic cocktail in the water every 2-3 days. These antibiotics reduced the bacterial load in the jelly coats by 98% throughout the course of embryogenesis, and thus potentially inhibited the production of bioactive secondary metabolites. Between days 8-14 of the experiment, antibiotic-treated eggs developed a thick biomass that resulted in embryo death. This biomass was primarily composed of the fungal pathogen, *Fusarium keratoplasticum*. To determine if there were host-derived compounds contributing to defense, eggs were also separated into their different components (intact eggs versus eggs with the jelly coat/outer capsule removed) and development was monitored with or without the presence of *F. keratoplasticum* conidia. Intact eggs developed normally in the presence of the fungal pathogen, but eggs with their jelly coat (and bacterial community) removed developed a fungal biomass and the embryos died. These experiments demonstrated that an active bacterial community in the jelly coat layer is necessary for normal *E.*

scolopes embryo development in the presence of a potential pathogen. Future studies will need to confirm if these findings are true in other cephalopods that possess an ANG, but due to the similarities in ANG/egg structures and bacterial communities, their function likely holds true.

The *Euprymna scolopes* model

The Hawaiian bobtail squid, *Euprymna scolopes,* is an emerging model for studying ANG symbioses, but it has been well-studied for over 30 years to examine beneficial-host microbe interactions in the light organ symbiosis (87)(88). The light organ of a juvenile bobtail squid is colonized by the bioluminescent bacterium, *Vibrio fischeri*, within hours of hatching (89). The light produced by *V. fischeri* is used by these nocturnal squid as a form of camouflage known as counterillumination. While the squid hunt at night, the light emitted by their light organ matches the down-welling moonlight, masking their silhouette from predators swimming below (90). *V. fischeri* makes up less than 0.1% of bacteria in the surrounding seawater (91), and it is the only species that can colonize the light organ (92). For this reason, *E. scolopes* has been used as a model to study the intricate process of colonization between one tissue type and one bacterium. Many questions have been investigated from both the host and symbiont perspective, such as how the host immune system can detect a symbiont from a background of non-symbiotic bacteria (93)(94), what bacterial genes are necessary for colonization (95), and how bacterial signals can influence the host's morphology (96)(97).

The small size of adult squid, rapid growth, and ease of maintenance in the lab make *E. scolopes* a good model for a number of research avenues. *E. scolopes* are endemic to the Hawaiian archipelago and adults are readily caught in shallow sand banks from two breeding populations found in Kaneohe Bay and Maunalua Bay on the island of Oahu (98). Wild adults

can be easily maintained in lab aquaria, where they will mate and females can lay several egg clutches over their lifetime, totaling hundreds of eggs (98)(99)(100). Although labor-intensive, juveniles can be reared to adulthood in the lab in 2-3 months, and their average life span has been reported to be 7-10 months (98)(99). Due to the horizontal acquisition of the light organ and ANG symbioses, juveniles can be raised with or without their symbionts by rearing in sterilized seawater (101). The recent availability of the *E. scolopes* genome and transcriptome (102) will likely lead to the development of new tools to make the host genetically tractable. Besides symbiosis research, *E. scolopes* has also been used as a model for the study of cephalopod development and the study of tissues that interact with light (100). The breadth of research that exists on the binary light organ symbiosis provides a foundation from which study of the ANG consortia can continue to draw parallels or reveal differences in how a host maintains and is affected by its microbiome.

Research Objectives

Defining the mechanisms responsible for a bacterial consortium's function in a host is a challenging process. The recent descriptions of the *E. scolopes* ANG community members and assignment of a defensive function have opened up many more questions regarding how the host acquires its consortium, its impacts on ANG development, and the bacterial metabolism necessary for egg defense. Challenges to answering these questions include the lack of cultured isolates from all community members, limited understanding of their metabolic potential, and little data for *in situ* bacterial metabolism. The core *E. scolopes* ANG community is composed of 52 operational taxonomic units (OTUs) present in 90% of samples, with 40 OTUs from the *Alphaproteobacteria* and 8 OTUs from the *Opitutae* (72). Prior to this thesis, all cultured

community members have been *Alphaproteobacteria*, the majority of which were *Leisingera* spp. within the *Rhodobacteraceae* family (103). Genomic analyses and/or biochemical assays revealed the potential for these isolates to produce a type VI secretion system, acyl homoserine lactones, siderophores, and gene clusters for the production of other secondary metabolites (103). *Leisingera* OTUs make up just 7 of the core OTUs, so even within the *Alphaproteobacteria* there is a greater diversity of species to be studied. Many of the other genera represented in the core OTUs have not been cultured. In particular, no isolates have been cultured from the second most dominant taxa, the *Verrucomicrobia*, which are members of a diverse group of bacteria found in many different ecosystems with very few cultured representatives (104).

The coupling of metagenomic and metatranscritpomic sequencing can reveal *in vivo* bacterial activity for systems where many culture-based experiments are still not feasible. Metatranscriptomics demonstrated the contributions of hindgut bacteria in wood digestion in higher termites (105) and bacterial roles in nitrogen fixation, amino acid biosynthesis, and wood digestion in lower termites (106). Metatranscriptomes of sponge symbionts demonstrated heterotrophic metabolisms, carbon fixation, coupled denitrification, and an interchange of amino acids between bacteria and host (107)(108). In the honey bee gut, metatranscriptomes showed how three major classes of bacteria (*Gammaproteobacteria, Bacilli, Actinobacteria*) are active in polysaccharide and polypeptide breakdown and generation of fermentation products for the host (109). In a defensive symbiosis, *Spiroplasma* transcripts had up-regulation of two putative toxin genes in flies exposed to a parasitic nematode (110). These “omic” methods were thus chosen to gain further insight into the bacterial metabolism in the ANG and egg jelly coats and how the bacteria interact with their host.

In chapter 2 and appendix 1 of my dissertation, the *in vitro* antibacterial and antifungal activity of ANG/JC isolates was examined. This activity was screened in a number of newly cultured isolates, including members of previously uncultured *E. scolopes* ANG taxa (*Flavobacteriia* and *Gammaproteobacteria*). Antimicrobial metabolite production was linked to two isolates in these studies, *Leisingera* sp. JC1 and *Pseudoalteromonas* sp. JC28. Chapter 3 discusses the metagenomic sequencing of the ANG bacterial community, providing metagenome assembled genomes (MAGs) from several taxa that have not yet been cultured, including *Verrucomicrobia*, *Alterythrobacter*, and *Mesorhizobium*. The community's potential for secondary metabolite production was also assessed, revealing biosynthetic gene clusters for bacteriocins, terpenes, polyketide synthases, and nonribosomal peptide synthetases, amongst others. Chapter 4 begins to examine the *in vivo* gene expression of the ANG bacterial community, focusing on both primary and secondary metabolism. Gene expression was compared between the ANG and egg jelly coat to determine an effect of transfer to the egg environment and decreasing population densities. Metatranscriptomes were also compared between control eggs and eggs challenged with a fungal pathogen to look for shifts in secondary metabolite production. Overall, antimicrobial activity was demonstrated in several members of the ANG/JC community. The metabolic potential of a *Verrucomicrobia* MAG was described, with particular attention to carbon utilization which may provide insights for culturing. Expression of secondary metabolite genes in the ANG and eggs demonstrated the potential for defensive bioactive compounds to be produced and provides a baseline of data to further explore the mechanisms of bacterial egg defense in *E. scolopes*.

Chapter 2 – Part 1

Accessory nidamental gland and egg jelly coat isolates inhibit marine bacteria *in vitro*: *Leisingera* sp. JC1, a bacterial isolate from Hawaiian bobtail squid eggs, produces indigoidine and differentially inhibits vibrios*

Significant Contributions

Significant contributions to this chapter were made by other researchers. Samantha Gromek, Jillian Garcia, and Marcy Balunas (Division of Medicinal Chemistry, Department of Pharmaceutical Sciences, University of Connecticut) created organic extracts of *Leisingera* sp. JC1, identified its pigment as indigoidine, performed extract assays, and wrote portions of the chapter. Matthew Fullmer and Johann Peter Gogarten performed taxonomic analysis, whole genome comparison, and wrote portions of the chapter.

*Citation:

Gromek SM*, Suria AM*, Fullmer MS, Garcia JL, Gogarten JP, Nyholm SV, and Balunas MJ (2016). *Leisingera* sp. JC1, a Bacterial Isolate from Hawaiian Bobtail Squid Eggs, Produces Indigoidine and Differentially Inhibits Vibrios. *Front. Microbiol*. 7:1342. doi: 10.3389/fmicb.2016.01342. *Authors contributed equally

frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 September 2016
doi: 10.3389/fmicb.2016.01342

Leisingera sp. JC1, a Bacterial Isolate from Hawaiian Bobtail Squid Eggs, Produces Indigoidine and Differentially Inhibits Vibrios

Samantha M. Gromek[1†], Andrea M. Suria[2†], Matthew S. Fullmer[2], Jillian L. Garcia[1], Johann Peter Gogarten[2,3], Spencer V. Nyholm[2] and Marcy J. Balunas[1*]*

[1] Division of Medicinal Chemistry, Department of Pharmaceutical Sciences, University of Connecticut, Storrs, CT, USA, [2] Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA, [3] Institute for Systems Genomics, University of Connecticut, Storrs, CT, USA

Female members of many cephalopod species house a bacterial consortium in the accessory nidamental gland (ANG), part of the reproductive system. These bacteria are deposited into eggs that are then laid in the environment where they must develop unprotected from predation, pathogens, and fouling. In this study, we characterized the genome and secondary metabolite production of *Leisingera* sp. JC1, a member of the roseobacter clade (*Rhodobacteraceae*) of *Alphaproteobacteria* isolated from the jelly coat of eggs from the Hawaiian bobtail squid, *Euprymna scolopes*. Whole genome sequencing and MLSA analysis revealed that *Leisingera* sp. JC1 falls within a group of roseobacters associated with squid ANGs. Genome and biochemical analyses revealed the potential for and production of a number of secondary metabolites, including siderophores and acyl-homoserine lactones involved with quorum sensing. The complete biosynthetic gene cluster for the pigment indigoidine was detected in the genome and mass spectrometry confirmed the production of this compound. Furthermore, we investigated the production of indigoidine under co-culture conditions with *Vibrio fischeri*, the light organ symbiont of *E. scolopes*, and with other vibrios. Finally, both *Leisingera* sp. JC1 and secondary metabolite extracts of this strain had differential antimicrobial activity against a number of marine vibrios, suggesting that *Leisingera* sp. JC1 may play a role in host defense against other marine bacteria either in the eggs and/or ANG. These data also suggest that indigoidine may be partially, but not wholly, responsible for the antimicrobial activity of this squid-associated bacterium.

Keywords: symbiosis, *Euprymna*, roseobacter, *Rhodobacteraceae*, indigoidine, *Leisingera*, DART-MS, secondary metabolite regulation

OPEN ACCESS

Edited by:
Russell T. Hill,
University of Maryland Center for Environmental Science, USA

Reviewed by:
Valerie McKenzie,
University of Colorado Boulder, USA
Anahit Penesyan,
Macquarie University, Australia

Correspondence:
Marcy J. Balunas
marcy.balunas@uconn.edu
Spencer V. Nyholm
spencer.nyholm@uconn.edu

†These authors have contributed equally to this work.

Specialty section:
This article was submitted to Microbial Symbioses, a section of the journal Frontiers in Microbiology

Received: *23 May 2016*
Accepted: *15 August 2016*
Published: *08 September 2016*

Citation:
Gromek SM, Suria AM, Fullmer MS, Garcia JL, Gogarten JP, Nyholm SV and Balunas MJ (2016) Leisingera sp. JC1, a Bacterial Isolate from Hawaiian Bobtail Squid Eggs, Produces Indigoidine and Differentially Inhibits Vibrios. Front. Microbiol. 7:1342. doi: 10.3389/fmicb.2016.01342

INTRODUCTION

It is becoming increasingly evident that many animals and plants use compounds produced by symbiotic bacteria for protection against pathogens and other fouling organisms (reviewed in Flórez et al., 2015). In marine and aquatic environments a number of invertebrates (including sponges, tunicates, bryozoans, and molluscs) host microorganisms that produce compounds used for such protection. These groups have served as an important source for studying defensive symbioses and for the discovery of novel bioactive natural products (see example in Schmidt and Donia, 2010).

Among molluscs, one common yet poorly understood animal–bacterial association occurs between members of squid and cuttlefish species and bacterial consortia that reside within a reproductive gland of female hosts called the accessory nidamental gland (ANG; Kaufman et al., 1998; Grigioni et al., 2000; Barbieri et al., 2001; Pichon et al., 2005; Collins et al., 2012). This organ harbors a dense consortium of bacteria housed in epithelium-lined tubules that are attached to the nidamental gland, the organ that secretes the jelly coat (JC) surrounding fertilized eggs. Bacteria from the ANG are deposited into the JC where they have been hypothesized to help protect developing eggs from fouling microorganisms, pathogens, and/or predation (Barbieri et al., 1997, 2001; Collins et al., 2012, 2015).

The Hawaiian bobtail squid, *Euprymna scolopes*, has been used as a model organism to study bacteria–host interactions, mainly due to the host's relationship with the bioluminescent bacterium *Vibrio fischeri* (McFall-Ngai, 2014). Recent studies have also focused on a second association found within the ANG of this species (Collins and Nyholm, 2011; Collins et al., 2012, 2015). These studies demonstrated that the ANG consortium in *E. scolopes* is dominated by members of the *Rhodobacteraceae* (roseobacters) within the *Alphaproteobacteria*, a common group of marine bacteria. A number of roseobacter-clade organisms are known to produce unique antimicrobial molecules and other secondary metabolites. For example, the antibiotic tropodithietic acid (TDA) and the algicidal roseobacticides are produced by *Phaeobacter* species and the antibacterial compound indigoidine is produced by *Leisingera* (formerly *Phaeobacter*) sp. Y4I (Geng et al., 2008; Seyedsayamdost et al., 2011; Cude et al., 2012). Most of these studies have focused on either free-living or plankton-associated roseobacters and the potential antimicrobial activity of the ANG strains has not been explored. A study that analyzed the genomes of 13 ANG roseobacter strains from *E. scolopes* did reveal the potential for secondary metabolite production (Collins et al., 2015) and *Gammaproteobacteria* from the ANG of another squid species have been shown to inhibit other bacteria (Barbieri et al., 1997).

In this study, we characterized the genome and secondary metabolite production of a new bacterial strain, *Leisingera* sp. JC1, isolated from the JC of *E. scolopes* squid eggs. Whole genome sequencing and biochemical analyses revealed the potential for and production of a number of secondary metabolites, including siderophores and acyl-homoserine lactones involved with quorum sensing. The complete indigoidine biosynthetic gene cluster was detected in the genome and mass spectrometry confirmed the production of this compound. Furthermore, we investigated the regulation of indigoidine under co-culture conditions with *V. fischeri*, the light organ symbiont. Finally, both *Leisingera* sp. JC1 and extracts from this strain exhibited differential antimicrobial activity against a number of marine vibrios, suggesting that indigoidine may be partially, but not wholly, responsible for the antimicrobial activity of this squid-associated bacterium.

MATERIALS AND METHODS

Bacterial Isolation

Hawaiian bobtail squid, *E. scolopes*, were obtained from sand flats in Oahu (Maunalua Bay, 21°16′51.42″ N, 157°43′33.07″ W), Hawaii and maintained in aquaria as previously described (Schleicher and Nyholm, 2011). Eggs laid in captivity from one adult female were collected, flash frozen on the 11th day of development, and stored at −80°C. Ten eggs were thawed for bacterial isolation and their outer capsules and embryos were removed and discarded with sterile forceps. The JCs were isolated, surface sterilized with 70% ethanol, and rinsed with filter-sterilized squid Ringers (FSSR, 530 mM NaCl, 25 mM $MgCl_2$, 10 mM $CaCl_2$, 20 mM HEPES, pH = 7.5). The 10 JCs were pooled and homogenized in FSSR, then serially diluted and plated on seawater tryptone (SWT) medium (5 g/L tryptone, 3 g/L yeast extract, 3 mL/L glycerol, 700 mL/L Instant Ocean sea salts, 15 g/L agar, 300 mL/L DI water). *Leisingera* sp. JC1 colonies appeared dark blue on this medium and were streaked to isolation.

Genomic Sequencing and Analysis

Genomic DNA was extracted using the MasterPure DNA Purification kit (Epicentre, Madison, WI, USA) from an overnight liquid culture of *Leisingera* sp. JC1 grown shaking at 30°C in SWT. DNA was quantified using a Qubit 2.0 fluorometer (Life Technologies, Agawam, MA, USA) and checked for quality on a 1% agarose gel and using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Agawam, MA, USA). A paired end library was prepared from 1 ng of genomic DNA using the Nextera XT DNA library kit (Illumina, Inc., San Diego, CA, USA) and quantified using the Qubit fluorometer and bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The library was sequenced on an Illumina MiSeq sequencer using 2 bp × 250 bp reads at the Microbial Analysis Resources and Services (MARS) facility at the University of Connecticut (Storrs, CT, USA).

Reads were trimmed using the CLC Genomic Workbench (Qiagen, Hilden, Germany) and a draft genome was assembled using the A5 assembler (Tritt et al., 2012). Coverage was determined by mapping trimmed reads to the draft genome assembly using CLC Genomic Workbench. The genome was annotated using the Rapid Annotation using Subsystem Technology (RAST, Aziz et al., 2008)[1] server and analyzed with the Antibiotic and Secondary Metabolite Analysis Shell 3.0 (antiSMASH, Weber et al., 2015)[2] for potential secondary metabolite biosynthesis gene clusters. The draft genome assembly has been deposited in DDBJ/EMBL/GenBank under accession LYUZ00000000. The version described in this paper is version LYUZ01000000.

[1] http://rast.nmpdr.org

[2] http://antismash.secondarymetabolites.org

Taxonomic Analysis and Whole Genome Comparison

Initial 16S identity suggested JC1 belonged to the genus *Leisingera* (data not shown). To validate this conclusion and to evaluate its relationship to the previously sequenced ANG isolates, a further taxonomic analysis was undertaken that used 17 previously described *Leisingera* genomes (Collins et al., 2015). A 33 gene multilocus sequence analysis was carried out following the methodology described in Collins et al. (2015). After generating alignments for each of the 33 genes using MUSCLE (Edgar, 2004), a concatenated alignment was generated using in-house python scripts. An optimal model of evolution was determined using the Akaike information criterion with correction for small sample size as implemented in jModelTest v2.1.4 (Darriba et al., 2012). The best-fitting model reported was GTR + Gamma estimation + Invariable site estimation. A maximum-likelihood (ML) phylogeny was generated from the concatenated multi-sequence alignment using PhyML v3.0_360-500M (Guindon et al., 2010). PhyML parameters consisted of GTR model, estimated p-invar, four substitution rate categories, estimated gamma distribution, sub-tree pruning and regrafting enabled with 100 bootstrap replicates. In addition to the maximum-likelihood analysis, a Bayesian inference analysis was also conducted using MrBayes v3.2.4 x64 (Ronquist et al., 2012). A mixed model with gamma estimation and invariable sites was used. The mixed model settled on a GTR submodel with only one parameter difference from the default GTR model with a posterior probability > 0.8. The standard GTR model accounted for the remainder of the model probability. The analysis used two cold chains with three heated chains each and ran for one million generations. After the run finished, convergence was assessed using average standard deviation of split frequencies of the cold chains, potential scale reduction factors of parameters, and minimum effective sample sizes of parameters. All criteria indicated the runs had converged.

Average nucleotide identity (ANI) was calculated using JSpecies 1.2.1 (Richter and Rosselló-Móra, 2009). The calculations were made using the MUMmer aligner with its default options. Contig files were generated for this analysis using the seqret function of the EMBOSS package (Rice et al., 2000). The reciprocal comparisons were averaged for reporting. Estimates of *in silico* DDH were made using the Genome-to-genome distance calculator 2.1 (Meier-Kolthoff et al., 2013) using the BLAST+ alignment method and the formula 2 algorithm outputs.

Select genomes were compared using the BLAST Ring Generator (BRIG) v1.0 (Alikhan et al., 2011). Default BLAST options were used. A whole genome alignment was generated using the Mauve program v2.3.1 (Darling et al., 2010). The progressiveMauve algorithm was used with default options.

Homoserine Lactone Detection

Homoserine lactone (HSL) production was detected using a well-diffusion assay with the HSL-sensing bacterium *Agrobacterium tumefaciens* NTL4 pZLR4 (Cha et al., 1998) as previously described (Ravn et al., 2001; Collins et al., 2015). In brief, *A. tumefaciens* NTL4 was grown in 3 mL of LB with 30 μg/mL gentamicin for 24 h at 30°C. This culture was used to inoculate 50 mL of AB minimal media with 0.5% casamino acids and 0.5% glucose (Chilton et al., 1974), and allowed to grow for another 24 h at 30°C. This culture was used to inoculate 100 mL of AB minimal media to which 1.2% agar had been added and a final concentration of 0.5% casamino acids, 0.5% glucose, and 75 μg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) was added after autoclaving. The inoculated molten agar was allowed to solidify in Petri dishes and wells were cut into the media using a sterile borer.

Leisingera sp. JC1 and *Leisingera* sp. ANG1 were grown overnight at 30°C in 3 mL of SWT broth with 30 μM $FeCl_3$, 0.5% glucose, and 0.5% casamino acids to induce HSL production. Cells were pelleted and the supernatant was collected and filtered through a 0.22 μm filter (Thermo Scientific, Agawam, MA, USA). Cell-free supernatant (60 μl) was added to the wells in the *A. tumefaciens* plates. The *N*-3-oxohexanoyl homoserine lactone standard was serially diluted and added to wells of an *A. tumefaciens* plate as a control and for semi-quantitative comparison. All plates were incubated at 28°C for 24 h before imaging.

Siderophore Detection

To qualitatively detect siderophore production, *Leisingera* sp. JC1 was plated in triplicate on chrome azurol S (CAS) indicator agar, modified for marine bacteria as previously described (Whistler and Ruby, 2003), and incubated at 28°C for 24 h before imaging. Sequestration of iron from CAS causes a color change from blue to orange, indicating siderophore production.

Detection of Indigoidine Biosynthesis Genes in JC1 Genomic DNA

To confirm the presence of indigoidine biosynthesis genes in JC1, genomic DNA was extracted and quantified as described for genomic sequencing above. Primers were designed (Supplemental Table S1) to amplify the *igiCDR* genes based on the draft genome assembly and using Primer3 software (Untergasser et al., 2012). PCR amplification was performed using the standard GoTaq Green Master Mix (Promega, Madison, WI, USA) protocol with 30 cycles and 55°C annealing temperature.

Leisingera sp. JC1 Large Scale Culture

Leisingera sp. JC1 was cultured for extraction using SWT media (as described above except without addition of glycerol, delineated hereafter as SWT_{ng}). A three step culturing process was employed to produce sufficient scale for secondary metabolite extraction, while ensuring that the bacterium was in late stationary phase for optimal production of secondary metabolites (Ruiz et al., 2010). First, small scale cultures were prepared by inoculating a JC1 colony into 5 mL of media in a 24 deep well plate, which was incubated for 3 days at room temperature while shaking at 200 rpm. Then, medium scale cultures were prepared by transferring 1.5 mL of the small scale cultures into 125 mL baffled flasks with 50 mL media, which were incubated for 3 days at room temperature while shaking at 125 rpm. Lastly, large scale cultures were prepared by transferring

15 mL of medium scale cultures into 1 L baffled flasks with 500 mL of media, which were incubated for 3 days at room temperature while shaking at 125 rpm.

Extraction of *Leisingera* sp. JC1

All extraction solvents were ACS grade and purchased from Sigma Aldrich (St. Louis, MO, USA).

Normal Extraction

Diaion HP20 resin (Supelco, Bellefonte, PA, USA) was pre-washed by sequentially rinsing resin with methanol and Millipore water (EMD Millipore, Billerica, MA, USA). Large scale JC1 cultures were sonicated to lyse cells prior to addition of pre-washed Diaion HP20 resin (50 g, 10% w/v), followed by incubation for 24 h at room temperature while shaking at 125 rpm. Bacterial culture and resin were then filtered using a coarse glass frit filter and washed with Millipore water to remove aqueous media components. The resin and bacterial culture were then sequentially extracted with methanol, dichloromethane, and acetone (2 × 150 mL). Organic portions were combined, extracted with ethyl acetate to remove residual aqueous material, and concentrated.

Indigoidine Enriched Extraction

Because indigoidine is poorly soluble in water and most organic solvents, a second extraction protocol was utilized to prepare an indigoidine enriched extract following modified literature procedures (Yu et al., 2013). Briefly, large scale cultures were sonicated to lyse cells and transferred to centrifuge tubes. Cells were then separated from supernatant by low-speed centrifugation (850 *g* × 5 min; Beckman Coulter Avanti J-E Centrifuge, Brea, CA, USA). Supernatant was transferred to new tubes and subjected to high-speed centrifugation (21,000 *g* × 10 min) to obtain an indigoidine enriched pellet. The pellet was washed with methanol, transferred to a microcentrifuge tube, dried under N_2 gas, and dissolved in dimethyl sulfoxide (DMSO).

Detection of Indigoidine Production by *Leisingera* sp. JC1 via LC–MS

All HPLC grade solvents and reagents were purchased from Sigma-Aldrich. LC–MS data were collected on an Agilent ESI single quadrupole mass spectrometer coupled to an Agilent 1260 HPLC system with a G1311 quaternary pump, G1322 degasser, and a G1315 diode array detector (Agilent Technologies, Santa Clara, CA). A gradient elution was used from 10% methanol in H_2O to 90% methanol in H_2O over 25 min using an Agilent Eclipse XDB-C_{18} RP-HPLC column (4.6 mm × 150 mm, 5 μm) and a flow rate of 1 mL/min. Indigoidine enriched extracts were prepared at 5 mg/mL in DMSO. Indigoidine eluted at retention time (t_R) 10.7 min in agreement with literature (Yu et al., 2013).

Zone of Inhibition Assays

To observe inhibition of vibrio strains and ANG isolate strains by *Leisingera* sp. JC1 (Supplementary Table S3), a zone of inhibition (ZOI) assay was used. The vibrio strains *V. anguillarum* 775, *V. parahaemolyticus* KNH1, *V. fischeri* ES114, *V. harveyi* B392, and *Photobacterium leiognathi* KNH6 were grown for 2.5 h (to stationary phase) at 30°C in YTSS (4 g/L tryptone, 2.5 g/L yeast extract, 15 g/L Instant Ocean sea salts) broth and then serially diluted from 10^7 to 10^4 CFU/mL in YTSS broth to observe density dependent inhibition. Each dilution was plated in triplicate on YTSS agar using a sterile swab to form a lawn. All ANG isolates tested were grown overnight (~4×10^8 CFU/mL) in SWT broth at 30°C and plated on SWT agar using a sterile swab to form a lawn. *Leisingera* sp. JC1 was grown overnight to a density of ~1×10^8 CFU/mL in SWT when testing with ANG isolates and in YTSS when testing with vibrio strains. This overnight broth of *Leisingera* sp. JC1 was spotted (10 μL) on the surface of each lawn in quadruplicate. All plates were incubated at 28°C for 24 h before imaging and ZOI measurements around the *Leisingera* sp. JC1 colonies. SWT or YTSS broth (10 μL) were spotted on each lawn as media controls, and 10 μL of the overnight culture of *Leisingera* sp. JC1 was spotted in quadruplicate on SWT or YTSS agar without any bacterial lawns as a growth control.

To quantify inhibition, an average of three ZOI diameters were measured and an average of three diameters of the JC1 colonies were measured using ImageJ (Schneider et al., 2012). Due to slight variations in JC1 colony size across trials, the measurements were normalized by subtracting the average JC1 colony diameter from the average ZOI diameter. To determine if differences in ZOIs across lawn densities per organism were statistically significant, one-way ANOVAs were performed. If the results of the one-way ANOVA indicated statistically significant differences, multiple comparisons *post hoc* Tukey tests were performed to determine which lawn densities were significantly different.

96-Well Liquid Assays

Leisingera sp. JC1 extracts were tested for antibacterial activity against *V. fischeri* ES114, *V. anguillarum* 775, and *V. parahaemolyticus* KNH1. High throughput assays with these bacterial strains were developed based on similar assays with natural product extracts and human pathogens (Zgoda and Porter, 2001), including obtaining CFU counts and growth curves for each of the vibrio strains as well as determining proper incubation times and temperatures and finding appropriate controls. These assays were performed in 96-well plates (Corning Costar, Corning, NY, USA) with SWT media and incubated at 28°C while shaking at 200 rpm. The bacterial inocula were prepared by adding select colonies into 5 mL of media and adjusted to OD_{600} 0.1 (approximately 1–2 × 10^8 CFU/mL as per Clinical and Laboratory Standards Institute, 2012). Colony forming unit (CFU) counts were manually confirmed to ensure accurate approximation for each vibrio strain.

Extracts were screened as previously described (Zgoda and Porter, 2001) with the following modifications. Briefly, master mix was prepared by addition of 1.6 mL adjusted vibrio inoculum, 7.84 mL sterile water, and 6.4 mL of SWT media. To each well, 198 μL of master mix was added with 2 μL of either positive control (chloramphenicol, final testing concentration 2.5 μg/mL), negative control (DMSO), or extract prepared in DMSO (screened at final concentration of 500 μg/mL; MIC performed using serial dilutions). Sterility control wells consisted

of 98 µL sterile water, 100 µL of SWT media, and 2 µL of DMSO. All controls and samples were tested in technical triplicates with experiments repeated a minimum of three times to confirm results. Plates were read at 600 nm every 2 h from 0 to 10 h with a final reading at 24 h using a Synergy H1 Hybrid Reader (Biotek, Winooski, VT, USA). Results are given as percent control activity (PCA) calculated in comparison with DMSO, the negative control.

Localization of Indigoidine Production by *Leisingera* sp. JC1 Using DART-MS

Direct analysis in real time-mass spectrometry (DART-MS) analysis was performed using a JEOL AccuTOF with DART ion source (IonSense, Inc., Saugus, MA, USA). High purity helium 5.0–6.0 grade (greater than 99.999% purity) was heated to 300°C and used for ionization. Five locations were selected on JC1 colonies in the presence or absence of *V. fischeri*, including (A) center of colony, (B) midpoint between center and edge of colony, (C) edge of colony, (D) ZOI (in the absence of *V. fischeri* sample was obtained from a point equidistant from colony edge), and (E) outside ZOI. At each location a sterile single use syringe needle (BD Medical, Franklin Lakes, NJ, USA) was placed in the sample and then placed between the DART ion source and the MS inlet. Positive ion MS data were obtained over a *m/z* range of 60–700 and relative percent abundance was obtained for the indigoidine ion. Standards were run after sampling each colony and mass spectral data were monitored in real time to ensure no residual indigoidine remained after each sample. DART-MS is only semi-quantitative due to the potential for differential ionization, suppression of ions, and/or changes in sample concentration in the DART ion source (Sanchez et al., 2011). Therefore, relative indigoidine ion abundance was used to generate heatmaps representing a gradient from less abundance (black) to more abundance (red).

Measurement of Indigoidine Production by *Leisingera* sp. JC1 in Co-culture

JC1 bacterial inoculum was prepared by adding JC1 colonies into 5 mL of SWT_{ng} media in a 24 deep well plate, incubated for 24 h at room temperature while shaking at 200 rpm. Bacterial inocula for the vibrios were prepared by adding bacterial colonies of each species separately into 5 mL of SWT_{ng} media in 24 deep well plates, incubated for 2 h at 28°C while shaking at 200 rpm. All bacterial inocula (JC1, *V. fischeri*, *V. anguillarum*, *V. parahaemolyticus*) were adjusted to OD_{600} 0.1 prior to use.

Co-cultures of JC1 with individual vibrios were prepared by adding 1 mL of adjusted JC1 inoculum to 10 mL SWT_{ng} media in 125 mL baffled flasks, incubated for 24 h at room temperature while shaking at 125 rpm, followed by addition of 200 µL of *V. fischeri*, *V. anguillarum*, or *V. parahaemolyticus*. After addition of the vibrio strain, co-cultures were incubated for an additional 24 h at room temperature while shaking at 125 rpm. Monocultures of JC1, *V. fischeri*, *V. anguillarum*, and *V. parahaemolyticus* were prepared by adding 1 mL of adjusted inoculum to 10 mL SWT_{ng} media in 125 mL baffled flasks, incubated for 48 h while shaking at 125 rpm.

All co-cultures and monocultures were extracted using the indigoidine enriched protocol described above. LC–MS data was obtained on the Agilent LC–MS system described above, using an isocratic method to ensure minimal baseline variation (10% acetonitrile in H_2O with 0.1% formic acid over 15 min at a flow rate of 1 mL/min with 20 µL injection volume). Extracts were prepared at 5 mg/mL in DMSO. Indigoidine was detected and quantitated via measurement of area under the curve at UV absorbance 299 nm and confirmed by MS.

RESULTS AND DISCUSSION

Genome Characteristics and General Metabolism

Taxonomic Placement of JC1

Leisingera sp. JC1 has a draft genome size of 5.19 Mb and GC content of 62.3% (**Table 1**), which is average for members of the roseobacter clade and similar to other squid-associated isolates (Collins et al., 2015). This larger genome size reflects the generalist lifestyle and ability to use diverse energy sources common of roseobacters (Newton et al., 2010). The *repABC* genes for plasmid replication are present as well as *tra* genes necessary for conjugative plasmid transfer, indicating the potential presence of extrachromosomal DNA. Further sequencing is necessary to confirm the number, size, and content of these putative plasmids.

Phylogenetic reconstruction methods (Bayesian and maximum-likelihood) used with the 33-gene concatenation returned identical topologies with overall strong statistical supports (**Figure 1**), placing JC1 close to the *Leisingera* taxa previously isolated from the ANG. Average nucleotide identity (ANI) and *in silico* DNA–DNA hybridization estimates (*is*DDH) support this placement. JC1 had higher ANI (90.5–91.7%) (Supplementary Table S2) and *is*DDH (38.8–44.6%) values with the ANG isolates than with any other *Leisingera* sp. These results also show that JC1 does not group with either *Leisingera* sp. ANG-M7 or the ANG1 group, but is still related to both (**Figure 1**; Supplementary Table S3), which is not unusual since other ANG isolates also fall outside the main ANG1 clade (Collins et al., 2015).

There are indications that JC1 may be more similar to *Leisingera* sp. ANG-M7 than to the *Leisingera* sp. ANG1 group. Both the ANI and *is*DDH values between JC1 and ANG-M7 are elevated in comparison to their values with the ANG1 group. There are no support statistics for ANI so it is uncertain if the 1.2% (JC1-M7 ANI versus JC1 compared to the ANG group) and 1.6% (M7-JC1 ANI versus M7 compared to the ANG group) higher values are significantly different. However, *is*DDH values are supported by 95% confidence intervals. The lower interval for JC1-M7 does not overlap with the upper interval for any comparison with a member of the ANG group, suggesting the *is*DDH values are significantly different. Additionally, the Bayesian inference found a small fraction of topologies in which the placements of JC1 and ANG-M7 were reversed, while the maximum-likelihood analysis found this occurrence in 33 of

TABLE 1 | Genome statistics of *Leisingera* sp. JC1.

Genome size (Mb)	Number of contigs	N_{50} (bp)	G + C content (%)	Number of genes	Missing genes* (% of total)	Fold coverage
5.19	168	123,213	62.3	5,074	54 (1.1)	37

*As predicted by the RAST server (Aziz et al., 2008).

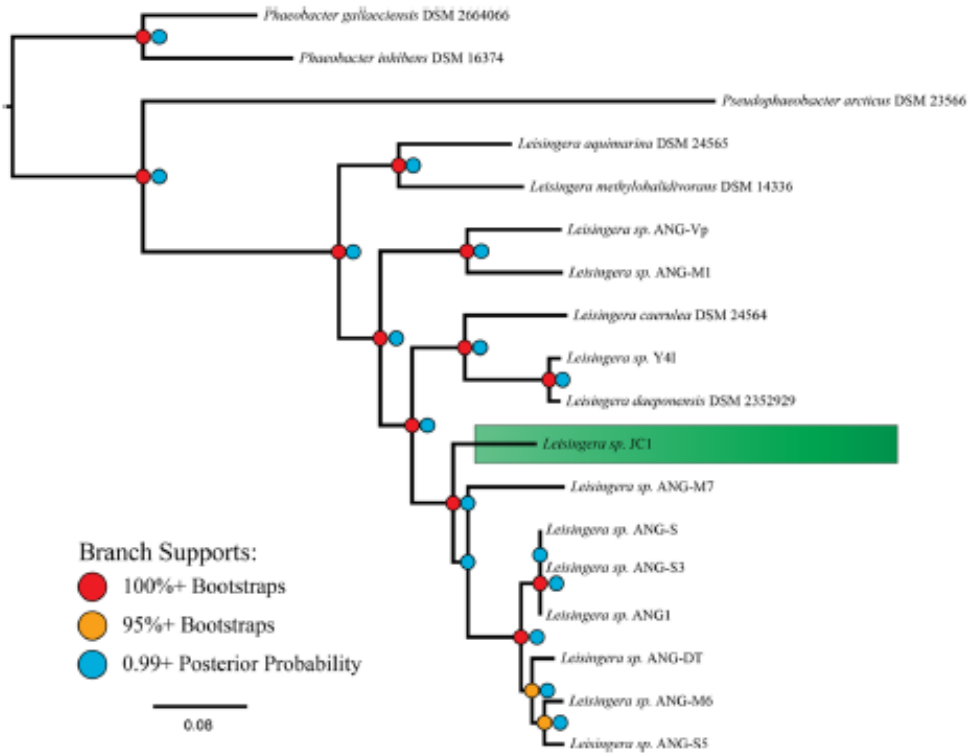


FIGURE 1 | MLSA analysis places *Leisingera* sp. JC1 in a sister group to other previously isolated ANG bacteria. Based on a comparison of 33 single-copy housekeeping genes, the egg jelly coat (JC) isolate JC1 is placed in the previously described roseobacter clade, "Clade 1" (Newton et al., 2010), along with seven previously isolated ANG bacteria (Collins et al., 2015). The phylogenetic tree reflects both Bayesian inference and maximum-likelihood methods with both posterior probability and bootstrap supports displayed.

100 bootstrap replicates. Overall, these analyses suggest that *Leisingera* sp. JC1 is distinct from, but related to the current ANG isolates.

Isolates having similar pigmentation to *Leisingera* sp. JC1 were cultured from other egg clutches, an ANG, and ovary from different females (data not shown). Among these, colonies with a similar dark blue morphology were isolated from the JCs of 1 and 23 day old eggs laid by different females. Similar colonies were isolated from the ANG of one of these females and the ovary of another female. Preliminary 16S sequencing placed two of these isolates in the genus *Leisingera* (data not shown), and further sequencing will reveal if these are the same strain as JC1. In addition, the production of the pigment indigoidine was confirmed by these additional strains (see below). These data suggest that *Leisingera* sp. JC1 and/or other indigoidine-producing strains may be selected for in the ANG/JC symbiosis.

Primary Metabolism

Leisingera sp. JC1 has a complete Entner-Duodoroff pathway and tricarboxylic acid cycle for metabolism of glucose. JC1 lacks any orthologs of phosphofructokinase, a major enzyme of glycolysis, but does contain a glucokinase and two distinct glucose-6-phosphate-1-dehydrogenases (GAPDHs). A glucose-6-phosphate-1-dehydrogenase (GPDH) is present, which catalyzes the first step of the alternative pathways for glucose metabolism, indicating that the Entner-Duodoroff pathway is probably used instead of glycolysis. *Leisingera* sp. JC1 only has the first two enzymes of the oxidative pentose phosphate pathway, but any 6-phosphate-gluconate produced can be further dehydrated by the Entner-Duodoroff pathway. Glycolate is a dissolved organic carbon often excreted by phytoplankton, and can be a carbon source for marine heterotrophic bacteria (Edenborn and Litchfield, 1985). *Leisingera* sp. JC1 is predicted

to oxidize glycolate to glyoxylate by a glycolate oxidase. JC1 has one system for glycerol uptake, the Ugp system, which can transport glycerol-3-phosphate against the concentration gradient. Sulfur oxidation genes are present, as well as a complete denitrification pathway with a copper-containing nitrite reductase. An assimilatory nitrate reductase is also present, which can convert nitrate to nitrite. An ammonia assimilation pathway is present with a ferrodoxin-dependent GOGAT, but no adenylyltransferase gene (GlnE) is present.

Transport

The high-affinity inorganic phosphate transport genes *pstABCS* and their regulatory genes *phoBUR* are present in JC1. The siderophore biosynthesis genes *asbAB* and *siderX456*, which encode high-affinity iron chelators, and the ferric iron ABC transporter, *pitADC*, are also present. JC1 has ABC transporters for dipeptides, oligopeptides, branched-chain amino acids, alkylphosphonate, and tungstate. The tripartite ATP-independent periplasmic (TRAP) transporter genes *dctMPQ* are present for unknown substrates, as well as the twin-arginine translocation (TAT) system genes, *tatABC*.

Leisingera sp. JC1 contains all 13 genes that encode the structural proteins essential for the Type VI Secretion System (T6SS) to function (Cianfanelli et al., 2016). The T6SS is a one-step mechanism for delivery of effectors across the Gram-negative outer membrane and membrane of the target cell, be it bacterial or eukaryotic. Widespread amongst the *Proteobacteria*, some T6SSs have been implicated in eukaryotic virulence (Pukatzki et al., 2006; Sana et al., 2012), but the majority are believed to play a role in bacterial competition (Hood et al., 2010; Schwarz et al., 2010). While it is possible for one T6SS system to affect both bacterial and eukaryotic targets (Jiang et al., 2014) it is believed that the system evolved for interactions with other bacteria, even in the case of intraspecific competition (Unterweger et al., 2014). Little work has been done, however, to investigate the role of T6SSs in beneficial host-symbiont relationships. Eleven of the 12 previously described ANG isolates also possess a T6SS (Collins et al., 2015), and it is possible that this system plays a role in interactions with other ANG or JC bacteria and/or the squid host. In the ANG, bacteria are partitioned into densely packed, epithelium-lined tubules, where each tubule is dominated by a particular taxon (Collins et al., 2012). These ANG/JC isolates may utilize the T6SS to outcompete other bacteria to establish colonization of a single tubule. While *Leisingera* sp. JC1 groups closely with other ANG isolates that also possess a T6SS (**Figure 1**, Collins et al., 2015), intraspecific effectors may facilitate competition between these strains, since ANG tubules are often highly pigmented with a single color (e.g., all dark blue matching the pigmentation of JC1 or all red-orange matching the pigmentation of several ANG isolates). Future studies will investigate the nature of JC1's T6SS effector proteins in the ANG symbiosis. There are numerous classes of evolved effector VgrG proteins, each with their own enzymatic function (reviewed in Durand et al., 2014). Understanding the number and type of effectors that JC1 can produce and deliver may help elucidate any role in the symbiosis.

Secondary Metabolite Biosynthesis

Analysis with the antibiotic and Secondary Metabolite Analysis Shell (antiSMASH, Weber et al., 2015) predicted several potential secondary metabolite biosynthesis gene clusters (Supplementary Table S4). These results included three separate siderophore clusters, one bacteriocin, one HSL, one type 1 polyketide synthase (T1 PKS), one other PKS (not type 1,2,3, or *trans*-AT), and two clusters classified as "other." Of these two "other" clusters, one contains the biosynthesis cluster for the known antimicrobial metabolite, indigoidine (Cude et al., 2012), while the other contains a previously described putative hybrid polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) gene cluster known to be conserved amongst roseobacters (Martens et al., 2007). This PKS/NRPS gene cluster encodes a polyketide synthase, glycosyl transferase, non-ribosomal peptide synthetase, and phosphopantetheinyl transferase, but the product of this cluster has not yet been identified. The top homologous gene cluster of the T1 PKS is 45% similar to a cluster in the ANG isolate, *Leisingera* sp. ANG-M7. While some roseobacters are capable of producing the novel secondary metabolite TDA (Bruhn et al., 2006, 2007; Geng et al., 2008), genes for synthesis of this molecule were not found nor was the molecule detected via LC–MS (data not shown).

Quorum Sensing

AntiSMASH predicted one *luxIR* homolog in *Leisingera* sp. JC1, flanked by an acyltransferase, crotonyl-CoA reductase, helicase, and oxidoreductase, similar to the previously published gene arrangement in bacterial isolates from the ANG (Collins et al., 2015). Production of HSLs by JC1 was confirmed in the *A. tumefaciens* NTL4 reporter assay, in which cell-free supernatant of a JC1 culture did induce β-galactosidase activity, indicating the presence of HSLs (**Figures 2A,B**). When compared to a dilution series of the *N*-3-oxohexanoyl HSL, JC1 produced a halo similar to that seen by 25 nM of HSL standard. The HSL production of JC1 was also slightly less than that of a closely related ANG isolate, *Leisingera* sp. ANG1.

Understanding the gene regulation by quorum sensing will be an important avenue of research for *Leisingera* sp. JC1 and the other *E. scolopes* ANG isolates due to the different habitats these bacteria experience. It is hypothesized that cephalopod ANGs are colonized via horizontal transmission from the environment (Kaufman et al., 1998), and potential symbionts must switch from living at very low cell densities in the seawater to very high cell densities in the ANG tubules (Collins et al., 2012). When ANG bacteria are deposited into the JC layers of eggs, these bacteria again experience a switch from the very high densities of the ANG to a lower density in the eggs. Due to this change in environments and cell densities, quorum sensing may play a role in gene regulation for ANG/egg JC bacteria.

Quorum sensing is also important in host-microbe interactions involving other roseobacters. For example, quorum sensing regulates motility and biofilm formation during host colonization in the sponge symbiont *Ruegeria* sp. KLH11 (Zan et al., 2012) and is necessary for colonization of the alga, *Ulva*

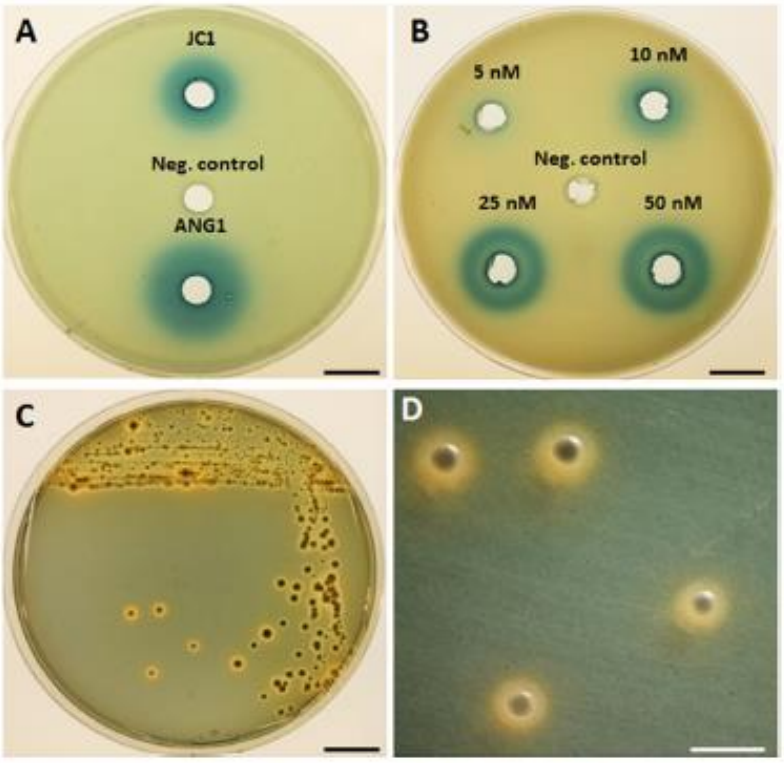


FIGURE 2 | Detection of homoserine lactone production and siderophore production by *Leisingera* sp. JC1. (A) Homoserine lactones were detected by β-galactosidase activity in cell-free supernatant of *Leisingera* sp. JC1 and compared with previously tested *Leisingera* sp. ANG1 (Collins et al., 2015); SWT broth was used as a negative control. **(B)** Dilution of *N*-3-oxohexanoyl homoserine lactone used as a positive control for the HSL assay; DMSO was the HSL standard solvent and was the negative control. **(C)** *Leisingera* sp. JC1 plated on CAS agar. Sequestration of iron changes the media from blue to orange, indicating siderophore production. **(D)** Magnified view of JC1 colonies from the plate in **(C)**, showing the orange halos in the media indicative of siderophore production. Scale bars, 1.5 cm **(A–C)**, 4 mm **(D)**.

australis by *Phaeobacter gallaeciensis* 2.10 (Rao et al., 2007). In other roseobacters, quorum sensing regulates secondary metabolite production, such as TDA in *Phaeobacter gallaeciensis* (Berger et al., 2011). In the indigoidine producing roseobacter, *Leisingera* sp. Y4I, there are two quorum sensing systems that regulate indigoidine production, *pgaIR* and *phaIR* (Cude et al., 2015). The JC1 *luxI* homolog has a 72% amino acid similarity to *pgaI* (RBY4I_1689) in Y4I, and the JC1 *luxR* homolog has an 81% amino acid similarity to *pgaR* (RBY4I_3631) in Y4I. The second set of *luxIR* homologs in Y4I, *phaIR* (RBY4I_3464 and RBY4I_1027), is not present in JC1. PgaI synthesizes the C8-HSL, produced by several proteobacteria, while PhaI synthesizes the $3OHC_{12:1}$-HSL, which may be species specific. JC1 lacking the *phaIR* system may reflect its divergence from *Leisingera* sp. Y4I. Further analyses will be needed to understand if indigoidine production in *Leisingera* sp. JC1 is regulated by quorum sensing.

Siderophore Production

Three separate siderophore biosynthesis gene clusters were detected in the genome, as described above, and production of iron chelators was confirmed by plating on CAS agar (**Figures 2C,D**). Appearance of an orange halo around colonies indicates that iron was sequestered from the chrome-azurol S dye in the media. Siderophores are high-affinity iron chelators, and can provide a growth advantage to cells in iron-limited environments, such as in seawater and in colonization of hosts. Although, the presence of siderophore biosynthesis genes in the genomes of currently sequenced roseobacter clade members is rare, 10 of the 12 previously sequenced *E. scolopes* ANG roseobacter symbionts did have the genes and/or demonstrate production of siderophores (Collins et al., 2015). Similar to the majority of the squid-associated roseobacter clade, the *Leisingera* sp. JC1 genome contains siderophore biosynthesis genes, indicating that siderophore production may play a role in the ANG symbiosis.

Indigoidine Biosynthesis Genes

The indigoidine biosynthesis gene cluster in *Leisingera* sp. JC1 contains all six biosynthesis genes previously described for *Leisingera* sp. Y4I (Cude et al., 2012) and shares a similar genome

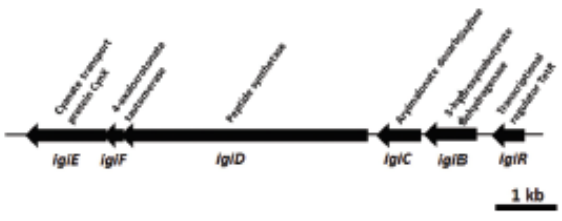


FIGURE 3 | *Leisingera* sp. JC1 indigoidine biosynthesis operon. Operon organization obtained from the published genome sequence (OBY26161.1, OBY26149.1, OBY26162.1, OBY26150.1, OBY26151.1, OBY26152.1). Arrows drawn to scale.

arrangement (**Figure 3**). To confirm the presence of individual members of the indigoidine biosynthesis gene cluster, primers were designed to three different components of the pathway, the non-ribosomal peptide synthetase (*igiD*), the transcriptional regulator (*igiR*), and one of the three indigoidine modification genes (*igiC*). The presence of these genes in JC1 genomic DNA was confirmed by PCR (Supplementary Figure S1).

Indigoidine biosynthesis genes have been detected in a diverse group of bacteria, including the *Actinobacteria* (*Streptomyces*), and *Alpha-*, *Beta-*, and *Gamma-proteobacteria*. The JC1 indigoidine biosynthesis operon shares the closest homology to the operon in *Leisingera* sp. Y4I, with 90–95% amino acid similarity for all gene products (**Table 2**). Other indigoidine biosynthesis operons share the non-ribosomal peptide synthetase, *igiD*, but many lack the same accessory genes required to modify indigoidine. When compared to other indigoidine producing strains, the *igiD* of JC1 is functionally homologous to other NRPS genes, sharing 49–53% amino acid similarity with *Vogesella indigofera*, *Streptomyces lavendulae*, and *Dickeya dadantii* 3937 (**Table 2**). A comparison with the genome of *Leisingera* sp. Y4I also confirmed that the indigoidine gene cluster is shared between these strains although absent from related ANG isolate *Leisingera* sp. M7 (Supplementary Figures S4 and S5).

Detection of Indigoidine Production by *Leisingera* sp. JC1

Because of the distinctive morphology and the genetic evidence for indigoidine biosynthesis, *Leisingera* sp. JC1 was cultured and extracted to obtain chemical evidence of indigoidine production. Using a three-step culture process, a deep blue liquid culture was obtained. However, upon extraction using a typical resin-based organic extraction protocol, most of the blue color was insoluble in organic solvents and little evidence of indigoidine production was observed via liquid chromatography-mass spectrometry (LC–MS, see **Figures 4D,E**), integrating to only 0.8% of the JC1 normal extract and indicative of negligible indigoidine extraction using this method. Therefore, an indigoidine enriched extraction protocol was utilized to pellet the insoluble indigoidine away from other media and cellular components followed by dissolving the sample in DMSO (Yu et al., 2013), resulting in an indigoidine enriched extract with 91.1% indigoidine. Analysis via LC–MS confirmed the presence of indigoidine (**Figure 4A**) in the indigoidine enriched extract (**Figures 4B,C**) with a peak eluting at 10.7 min with an $[M\text{-}H]^-$ of 247.0, consistent with the molecular weight and fragmentation pattern of indigoidine (248.2 g/mol) and in agreement with literature precedent (Yu et al., 2013). In addition, indigoidine was detected in two other JC and ANG isolates that exhibited a similar dark blue coloration in culture (data not shown).

Antibacterial Activity of *Leisingera* sp. JC1

Zone of Inhibition

Zone of inhibition assays were performed to test the ability of *Leisingera* sp. JC1 to inhibit other marine bacteria, both free-living and symbiotic (Supplementary Table S3). JC1 was tested against the *E. scolopes* light organ symbiont, *V. fischeri* ES114; another bioluminescent member of the *Vibrionaceae*, *P. leiognathi* KNH6, isolated from Hawaiian seawater; *V. harveyi* B392; *V. parahaemolyticus* KNH1 and *V. anguillarum* 775. These bacteria were plated at lawn densities from 10^4 to 10^7 CFU/mL to test the efficacy of possible inhibition at varying densities

TABLE 2 | Comparison of indigoidine biosynthesis operon in *Leisingera* sp. JC1 to other indigoidine producing strains.

Gene	Annotation	% Amino acid identity to *Leisingera* sp. Y4I operon	% Amino acid identity to *Vogesella indigofera* operon	% Amino acid identity to *Streptomyces lavendulae* operon	% Amino acid identity to *Dickeya dadantii* 3937 operon
igiE	Cyanate transport protein, CynX	96	58	NA	NA
igiF	4-oxalocrotonate tautomerase	91	NA*	NA	NA
igiD	Peptide synthetase	91	53	50	49
igiC	Arylmalonate decarboxylase	96	56	NA	NA
igiB	Hydroxyisobutyrate dehydrogenase	93	51	NA	NA
igiR	Transcriptional regulator, TetR	90	42	NA	NA

*NA, not applicable, no homolog of the gene present in that organism.

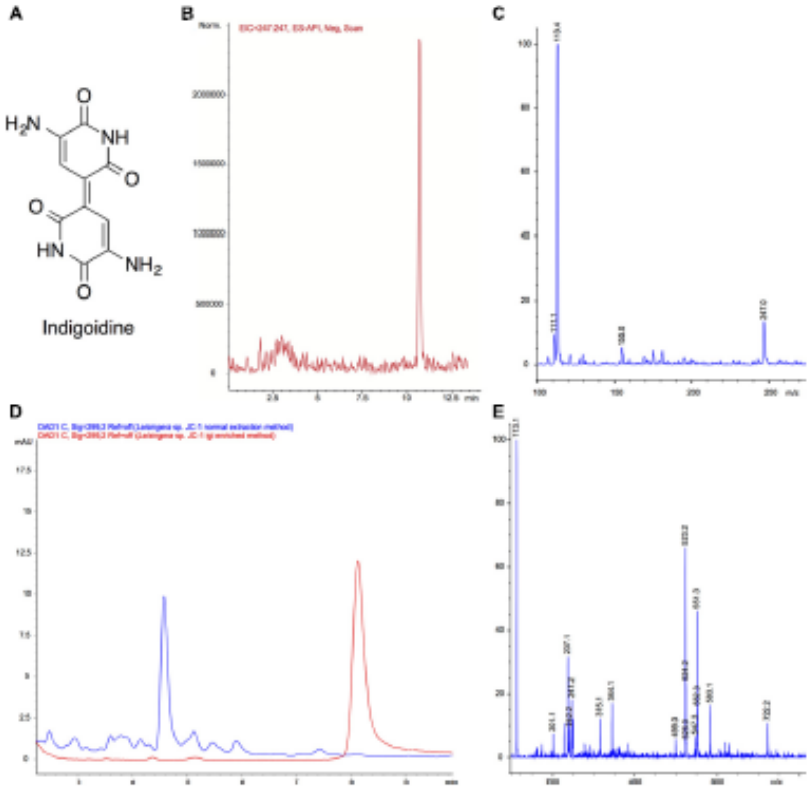


FIGURE 4 | Mass spectral confirmation of indigoidine production by *Leisingera* sp. JC1 and comparison of normal and IgI enriched extracts. **(A)** Structure of indigoidine. **(B)** Extracted ion chromatogram (EIC) of indigoidine to confirm presence in extract ($[M-H]^-$ 247.0) using a previously reported LC-MS solvent gradient (Yu et al., 2013). **(C)** Negative ionization mass spectrum of peak at 10.7 min. The $[M-H]^-$ peak of 247.0 is consistent with a molecular formula of $C_{10}H_7N_4O_4$, confirming the presence of indigoidine (248.19 g/mol). **(D)** LC–MS UV chromatogram overlays (299 nm) of normal extract (blue) with indigoidine (IgI) enriched extract (red). The peak eluting at t_R 8.2 min in the IgI enriched extract was confirmed to be indigoidine with a purity of 91.1% (t_R different due to use of isocratic solvent system to minimize baseline variation). Although, a small peak was present at t_R 8.2 min in the normal extract, this integrates to only 0.8% indicating that the *Leisingera* sp. JC1 normal extract contains negligible amounts of indigoidine. **(E)** Negative ionization mass spectrum of peak at t_R 8.2 min for normal extraction method. Although there was an ion corresponding to $[M-H]^-$ of 247.2, this was not the most dominant ion and the more quantitative UV analysis in D shows only a very small peak, indicative of negligible indigoidine in the JC1 normal extract.

which more closely reflect biologically relevant concentrations. Overall, *Leisingera* sp. JC1 differentially inhibited the five vibrios tested (**Figure 5**). For two of the strains tested, *V. fischeri* ($F_{3,76} = 12.63$, $P < 0.0001$) and *P. leiognathi* ($F_{3,60} = 137.5$, $P < 0.0001$), JC1 showed significantly greater inhibition at lower lawn densities (**Figure 5A**; Supplementary Figure S2). When measured ZOIs were normalized for variations in JC1 colony diameter, there was an average 4.2 mm ZOI at 10^4 CFU/mL of *V. fischeri*, while at the 10^7 CFU/mL density, there was a 2.3 mm ZOI (Supplementary Figures S2I–L). A multiple comparisons *post hoc* Tukey test determined that the ZOI for the 10^4–10^6 CFU/mL lawn densities of *V. fischeri* were significantly greater than the ZOI at the 10^7 CFU/mL density. The change in ZOI with test strain lawn density was most apparent for *P. leiognathi*, where the average ZOI at the 10^4–10^5 CFU/mL lawn densities ranged from 2.4 to 2.9 mm, and then dropped to 0 mm at the 10^6–10^7 CFU/mL densities (**Figures 5A,D**; Supplementary Figures S2A–D). A multiple comparisons *post hoc* Tukey test showed that the ZOI at 10^4 CFU/mL of *P. leiognathi* was significantly different from the ZOI at 10^5 CFU/mL, and that both ZOIs at 10^4 and 10^5 CFU/mL were significantly different from the 10^6–10^7 CFU/mL results. *Leisingera* sp. JC1 showed a trend toward inhibition of *V. anguillarum* with ZOIs ranging from 2.7 to 3.3 mm (**Figures 5A,E**; Supplementary Figures S2E–H) although, a one-way ANOVA determined that the ZOIs were not statistically different ($F_{3,60} = 2.553$, $P = 0.0639$). No inhibition was observed when JC1 was tested against *V. parahaemolyticus* or *V. harveyi* at any lawn density (**Figures 5A,F,G**; Supplementary Figures S2M–P).

Leisingera sp. JC1 was also tested in a ZOI assay against the 12 previously described ANG isolates (Collins et al., 2015) and one additional ANG isolate, *Muricauda* sp. ANG21. All ANG isolates were only tested at a lawn density of approximately 10^8 CFU/mL. Inhibition was observed against *Ruegeria* sp. ANG-S4,

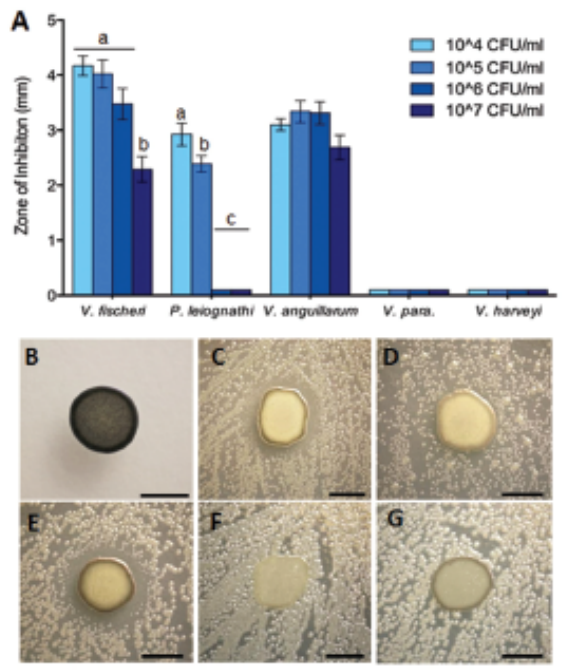


FIGURE 5 | *Leisingera* sp. JC1 differentially inhibits pathogenic and non-pathogenic vibrio species. (A) Zones of inhibition around JC1 colonies were measured with increasing lawn densities of each vibrio species. *Leisingera* sp. JC1 differentially inhibited five vibrio strains. Significantly greater inhibition of *Vibrio fischeri* was seen between the 10^4–10^6 CFU/ml and the 10^7 CFU/ml lawn density. Significantly greater inhibition of *Photobacterium leignathi* was seen between the 10^4 and 10^5 CFU/ml lawn densities, as well as between the 10^4–10^5 and the 10^6–10^7 CFU/ml densities. JC1 inhibited *Vibrio anguillarum* uniformly across all lawn densities and did not inhibit *Vibrio parahaemolyticus* or *Vibrio harveyi*. Letters indicate significantly different groups based on results from *post hoc* Tukey tests (see Results and Discussion). Representative 24 h images of **(B)** JC1 plated alone on YTSS agar and with 10^6 CFU/ml of **(C)** *V. fischeri*, **(D)** *P. leignathi*, **(E)** *V. anguillarum*, **(F)** *V. parahaemolyticus*, and **(G)** *V. harveyi*. Scale bars: 5 mm **(B–G)**.

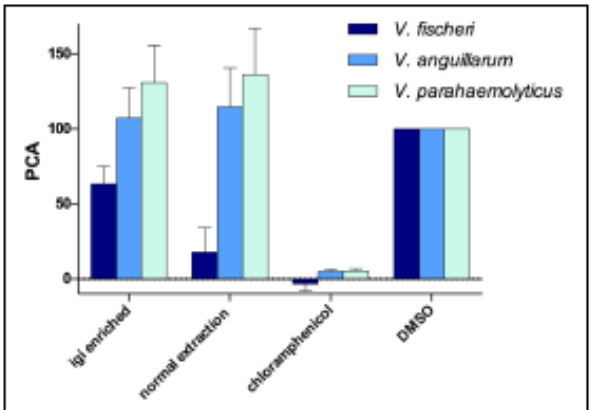


FIGURE 6 | *Leisingera* sp. JC1 extracts inhibit *V. fischeri* in 96-well liquid assays. JC1 cultures were grown at large scale and extracts were prepared using either an indigoidine enriched procedure (IgI enriched) or a normal extraction method. All extracts were screened at 500 μg/mL against *V. fischeri*, *V. anguillarum*, and *V. parahaemolyticus*. The JC1 normal extract strongly inhibited growth of *V. fischeri* (PCA 17.9 ± 9.3) with less activity of the JC1 IgI enriched extract against *V. fischeri* (PCA 63.3 ± 6.7). No inhibition of *V. anguillarum* or *V. parahaemolyticus* was observed. Data shown are average PCA values from three experiments (PCA, percent control activity, as compared with DMSO negative control).

with an average ZOI of 6.3 mm (±0.7) and against *Muricauda* sp. ANG21, with an average ZOI of 5.9 mm (±0.6; Supplementary Figure S3). *Leisingera* sp. JC1 was not able to inhibit any of the other *Leisingera* spp. previously isolated from ANGs. Since partitioning between bacterial taxa is observed in the ANG tubules some activity against other ANG isolates may contribute to competition between strains during colonization (Collins et al., 2012).

96-Well Liquid Assay

Both the normal and indigoidine enriched JC1 extracts were screened for activity using 96-well plate liquid assays with several of the vibrios tested above, including *V. fischeri* ES114, *V. anguillarum* 775, and *V. parahaemolyticus* KNH1 (**Figure 6**). Both extracts were initially tested at 500 μg/mL with minimum inhibitory concentrations (MICs) determined for active samples. The JC1 normal extract was found to strongly inhibit growth of *V. fischeri* with a PCA value of 17.9 ± 9.3 during screening and was determined to have a MIC of 250 μg/mL. The JC1 indigoidine enriched extract also exhibited moderate inhibition of *V. fischeri* with a PCA of 63.3 ± 6.7.

In contrast to the ZOI data above, no inhibition was observed for either extract when tested against *V. anguillarum*, potentially due to differences between the activity of indigoidine in agar versus liquid assays, as seen with *Leisingera* sp. Y41 and hypothesized to result from changes in the redox state of indigoidine (Cude et al., 2012). These results may also be attributed to differences in the chemical composition between extracts and the bacteria *in situ* (e.g., aqueous soluble metabolites are generally excluded from the extraction protocols used in this study). Neither JC1 extract inhibited *V. parahaemolyticus*, in agreement with the ZOI data above.

Previous studies with a mutant of *Leisingera* sp. Y4I that did not produce indigoidine suggested that production of the compound is required for inhibition of *V. fischeri* (Cude et al., 2012). However, with the more potent inhibition of *V. fischeri* seen in the JC1 normal extract versus the indigoidine enriched extract in this study (**Figure 6**), indigoidine production does not seem to be the only mechanism of inhibition for *Leisingera* sp. JC1. Given that the JC1 normal extract contains only minimal amounts of indigoidine (0.8% as discussed above), the bacterium may be utilizing other secondary metabolites in conjunction with indigoidine for chemical defense. The JC1 genome includes several other secondary metabolite biosynthetic gene clusters for HSL, siderophore, bacteriocin, PKS, and PKS/NRPS production and thus *Leisingera* sp. JC1 likely utilizes one or more of the

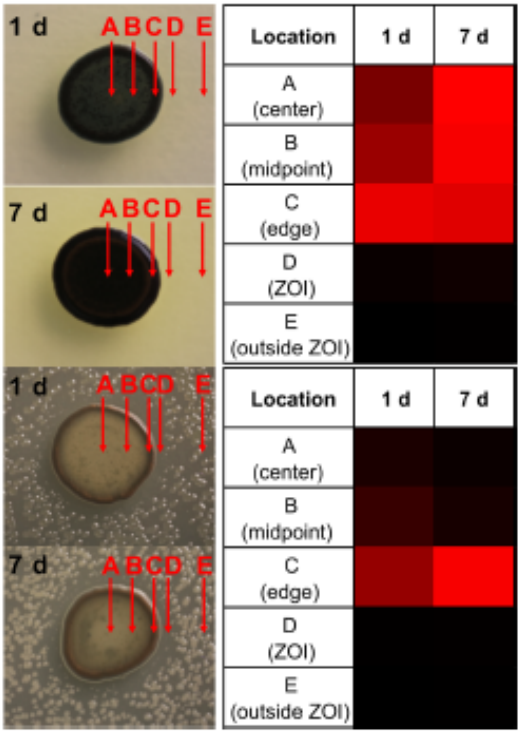


FIGURE 7 | *Leisingera* sp. JC1 localizes indigoidine production to outer edge of colony in the presence of *V. fischeri*. Five locations (A–E) were analyzed for the presence of indigoidine via direct analysis in real time-mass spectrometry (DART-MS) over 7 days. (A) In the absence of *V. fischeri*, indigoidine production was more uniformly distributed across the colony. (B) In the presence of *V. fischeri*, indigoidine production was primarily localized to outer edge of colony (location C) whereas minimal indigoidine was detected in the center of the colony (location A). Very little indigoidine was detected in the ZOI or outside ZOI in the presence or absence of *V. fischeri*. Heatmaps represent a gradient of indigoidine ion abundance from less abundance (black, 0%) to more abundance (red, 100%).

compounds encoded by these pathways for chemical defense, in addition to the defensive capabilities attributed to indigoidine. Creating an indigoidine mutant of *Leisingera* sp. JC1 will help test this hypothesis, in conjunction with identification of additional metabolite(s) responsible for JC1 antimicrobial activity.

Localization of Indigoidine Production by *Leisingera* sp. JC1

While performing ZOI assays, there was a dramatic change in colony pigmentation of *Leisingera* sp. JC1 when grown alone (**Figure 5B**) as compared to growth under challenge with various vibrio strains (**Figures 5C–G**). Deep blue pigment production was observed uniformly when JC1 was grown in monoculture and appeared to localize to the outer edges of the colonies when presented with vibrio strains. Direct analysis in real time-mass spectrometry (DART-MS) is an ambient ionization technique in which samples can be analyzed without sample preparation or extraction (Sanchez et al., 2011). DART-MS was utilized to chemically confirm the visual observations of localization of indigoidine production of JC1 in monoculture and co-culture with *V. fischeri* over the course of 7 days (**Figure 7**). Five locations were selected on each colony including center (A), midpoint (B), edge (C), ZOI (D), and outside ZOI (E). In the absence of *V. fischeri*, indigoidine was uniformly produced throughout JC1 colonies (locations A–C). However, in the presence of *V. fischeri* there was little to no indigoidine production in the center or midpoints of JC1 colonies, but intense indigoidine detected around colony edges (**Figure 7**, location C only). Indigoidine was only minimally detected in the ZOI or outside the ZOI for either monoculture or co-culture. Trends in the localization of indigoidine production were even more apparent upon measurement after 7 days.

There are several examples of pigment production being induced when in the presence of other bacteria, such as in *Staphylococcus aureus* when co-cultured with *Pseudomonas aeruginosa* (Antonic et al., 2013) or production of a red pigment by *Streptomyces lividans* TK23 when co-cultured with *Tsukamurella pulmonis* TP-B0596 (Onaka et al., 2011). Pigment production can also be induced under other stress response conditions, such as protection from UV radiation (Tong and Lighthart, 1997). Pigment production has also been tied to photosynthesis (Orf and Blankenship, 2013), however, *Leisingera* sp. JC1 lacks genes associated with photosynthesis or carbon fixation (data not shown).

When grown alone, *Leisingera* sp. JC1 exhibited a uniform blue–black pigmentation across the colony which was confirmed by mass spectrometry to be essentially uniform production of indigoidine. Secondary metabolite biosynthesis is an energy intensive endeavor and production of antimicrobial compounds would typically be thought to be reserved for defensive situations. Since indigoidine is produced throughout the colony when in monoculture, and given its relatively moderate antibacterial activity as suggested by assays with the indigoidine enriched extract, it is also possible that indigoidine serves multiple functions for *Leisingera* sp. JC1. However, *Leisingera* sp. JC1 localized indigoidine production to the outer edges of the colony when co-cultured with *V. fischeri* and other vibrios. If utilized as a defensive compound, indigoidine may be localized to points of direct interaction with other microorganisms. Secondary metabolite production can be localized to susceptible locations such as in plants, sponges, and other sessile terrestrial and marine organisms (Amsler et al., 2001; Furrow et al., 2003; Van Dyck et al., 2010). The role of *Leisingera* sp. JC1 has yet to be examined directly in the ANG symbiosis but localized production of indigoidine or other secondary metabolites may play a role in egg defense or inhibition of other bacteria from colonizing the ANG (see conclusions below).

Regulation of Indigoidine Production by *Leisingera* sp. JC1

After observing localized production of indigoidine when grown on solid media with *V. fischeri*, additional co-culture experiments were undertaken in liquid media using several

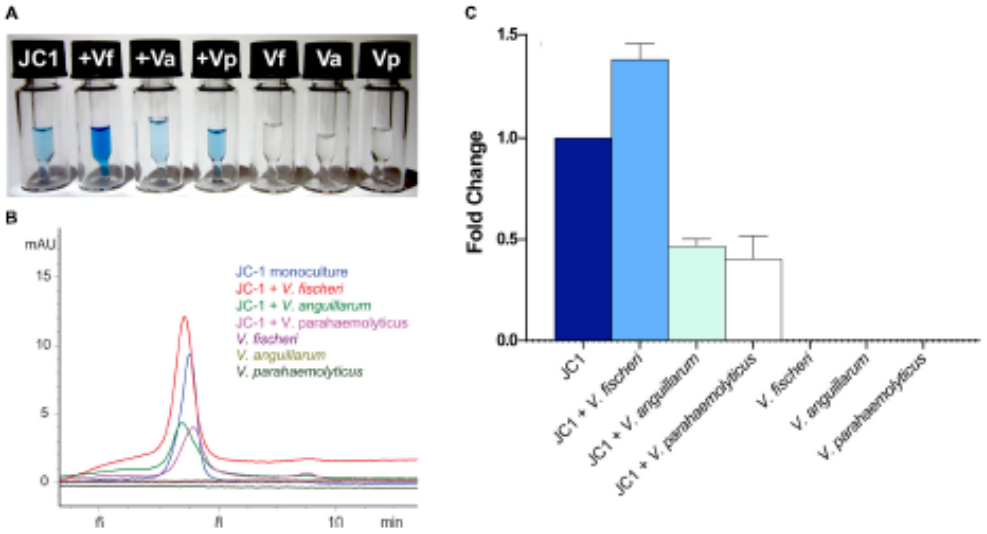


FIGURE 8 | *Leisingera* sp. JC1 upregulates indigoidine production in the presence of *V. fischeri*. JC1 was cultured alone and in the presence of *V. fischeri*, *V. anguillarum*, and *V. parahaemolyticus* (each vibrio was also monocultured). **(A)** Extracts were prepared in DMSO at 5 mg/mL for LC–MS (from left to right): JC1 monoculture, JC1 + *V. fischeri*, JC1 + *V. anguillarum*, JC1 + *V. parahaemolyticus*, *V. fischeri* monoculture, *V. anguillarum* monoculture, and *V. parahaemolyticus* monoculture. As shown, the highest pigment production was observed when JC1 was co-cultured with *V. fischeri*. **(B)** LC–MS UV chromatograms (299 nm) for all extracts. JC1 indigoidine production was upregulated when co-cultured with *V. fischeri* (red). Down regulation was observed when co-cultured with *V. anguillarum* or *V. parahaemolyticus* (green and pink, respectively). **(C)** Following extraction, indigoidine production was measured using HPLC and quantitated via measurement of the area under the curve for UV absorption at 299 nm. Fold change was calculated as compared with JC1 monoculture.

of the vibrios from the antimicrobial assays above. *Leisingera* sp. JC1 was grown in monoculture and in the presence of *V. fischeri*, *V. anguillarum*, and *V. parahaemolyticus*, followed by extraction and measurement of indigoidine production (**Figure 8**). Monocultures of all three vibrios were also grown and extracted as controls. Addition of *V. fischeri* to established cultures of *Leisingera* sp. JC1 resulted in a 1.38 fold increase in indigoidine. Co-cultures with *V. anguillarum* and *V. parahaemolyticus* resulted in a decrease in indigoidine production of approximately 0.5 fold for both organisms. Vibrio monocultures confirmed that these species do not produce indigoidine. Changes in indigoidine production were also visually evident with darker, more intense blue observed for extracts cultured with *V. fischeri* in comparison with JC1 monoculture, as well as lighter blue extracts observed for *V. anguillarum* and *V. parahaemolyticus* co-cultures (**Figure 8A**).

The increase in indigoidine production of JC1 with *V. fischeri* is consistent with the antibacterial activity observed for *Leisingera* sp. JC1 on both solid and liquid media (**Figures 5** and **6**), strengthening the hypothesis that indigoidine may play a protective role in association with *E. scolopes*. In addition, the downregulation of production with *V. anguillarum* and *V. parahaemolyticus* also supports the liquid culture bioassay data (**Figure 6**). The differential antimicrobial activity and indigoidine production between the three vibrios may be due to the purported role of the ANG and JC bacteria in the host. The ability of *Leisingera* sp. JC1 to inhibit *V. fischeri* may be related to the fact that the ANG is located directly posterior to the light organ, which harbors high densities of the sole symbiont, *V. fischeri* (McFall-Ngai, 2014). Each day 95% of viable *V. fischeri* cells in the light organ are expelled directly into the mantle cavity of the host as part of the regulatory mechanisms of that association (Boettcher et al., 1996; Nyholm and McFall-Ngai, 1998). A study from another squid, *Doryteuthis pealeii* (Kaufman et al., 1998) suggests that ANG bacteria are environmentally transmitted during development. Given that *V. fischeri* is not detected in the ANG (Collins et al., 2012), the inhibitory effect of *Leisingera* sp. JC1 and other ANG isolates may prevent *V. fischeri* and other vibrios from colonizing the ANG and thus help shape the consortium during development. Alternatively, inhibition against vibrios may play a role in egg defense since eggs are exposed to seawater for approximately three weeks and vibrios are known to be common members of the bacterioplankton.

CONCLUSIONS

Genome analyses confirm that *Leisingera* sp. JC1 is part of the squid-associated roseobacter clade. Both *in silico* and *in vitro* analyses confirmed the secondary metabolite potential and production of siderophores, acyl-homoserine lactones associated with quorum sensing, and the pigment indigoidine. *Leisingera* sp. JC1 and its extracts had inhibitory activity against a variety of marine bacteria including the light organ symbiont *V. fischeri*. Furthermore, JC1 challenged with *V. fischeri* led to increased localized

production of indigoidine as well as an increased production of indigoidine when co-cultured in liquid media. Taken together these results suggest that *Leisingera* sp. JC1 may play a protective role in egg defense and/or in shaping the microbial community of the ANG. The importance of defensive symbioses in nature is becoming increasingly more evident (Flórez et al., 2015). A number of both terrestrial and marine organisms use novel secondary metabolites produced by bacteria toward defense from potential pathogens and fouling microorganisms. Since roseobacters have been found in the ANGs of a number of cephalopods from diverse marine environments (Kaufman et al., 1998; Grigioni et al., 2000; Barbieri et al., 2001; Pichon et al., 2005; Collins et al., 2012) there may be a conserved function of this group in this symbiosis. Further studies from this group may reveal novel compounds that are important for the biology of these associations and that exhibit antimicrobial activity.

AUTHOR CONTRIBUTIONS

MB, SN, SG, and AS conceptualized and designed research; SG, AS, MF, and JLG conducted experiments; MB, SN, SG, AS, MF, and JPG analyzed data and wrote the paper.

FUNDING

This research was funded by NSF IOS-1557914 to SN and MB, University of Connecticut Office of the Vice President for Research to SN, and the University of Connecticut Outstanding Multicultural Scholars Program to AS.

ACKNOWLEDGMENTS

The authors would like to thank Anne A. Sung for assistance with CFU counts, Alison Buchan for donation of *Leisingera* sp. Y4I, Allison H. Kerwin for helpful comments, and Kewalo Marine Laboratory of the University of Hawaii for assistance with animal collections.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.01342

REFERENCES

Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402

Amsler, C. D., McClintock, J. B., and Baker, B. J. (2001). Secondary metabolites as mediators of trophic interactions among Antarctic marine organisms. *Am. Zool.* 41, 17–26.

Antonic, V., Stojadinovic, A., Zhang, B., Izadjoo, M. J., and Alavi, M. (2013). *Pseudomonas* aeruginosa induces pigment production and enhances virulence in a white phenotypic variant of *Staphylococcus aureus*. *Infect. Drug Resist.* 6, 175–186. doi: 10.2147/IDR.S49039

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Barbieri, E., Barry, K., Child, A., and Wainwright, N. (1997). Antimicrobial activity in the microbial community of the accessory nidamental gland and egg cases of Loligo pealei (Cephalopoda: Loliginidae). *Biol. Bull.* 193, 275–276.

Barbieri, E., Paster, B., Hughes, D., Zurek, L., Moser, D., Teske, A., et al. (2001). Phylogenetic characterization of epibiotic bacteria in the accessory nidamental gland and egg capsules of the squid Loligo pealei (Cephalopoda:Loliginidae). *Environ. Microbiol.* 3, 151–167. doi: 10.1046/j.1462-2920.2001.00172.x

Berger, M., Neumann, A., Schulz, S., Simon, M., and Brinkhoff, T. (2011). Tropodithietic acid production in *Phaeobacter gallaeciensis* is regulated by N-acyl homoserine lactone-mediated quorum sensing. *J. Bacteriol.* 193, 6576–6585. doi: 10.1128/JB.05818-11

Boettcher, K. J., Ruby, E. G., and McFall-Ngai, M. J. (1996). Bioluminescence in the symbiotic squid *Euprymna scolopes* is controlled by a daily biological rhythm. *J. Comp. Physiol.* 179, 65–73. doi: 10.1007/BF00193435

Bruhn, J. B., Gram, L., and Belas, R. (2007). Production of antibacterial compounds and biofilm formation by *Roseobacter* species are influenced by culture conditions. *Appl. Environ. Microbiol.* 73, 442–450. doi: 10.1128/AEM.02238-06

Bruhn, J. B., Haagensen, J. A. J., Bagge-Ravn, D., and Gram, L. (2006). Culture conditions of *Roseobacter* strain 27-4 affect its attachment and biofilm formation as quantified by real-time PCR. *Appl. Environ. Microbiol.* 72, 3011–3015. doi: 10.1128/AEM.72.4.3011-3015.2006

Cha, C., Gao, P., Chen, Y. C., Shaw, P. D., and Farrand, S. K. (1998). Production of acyl-homoserine lactone quorum-sensing signals by gram-negative plant-associated bacteria. *Mol. Plant Microbe Interact.* 11, 1119–1129. doi: 10.1094/MPMI.1998.11.11.1119

Chilton, M. D., Currier, T. C., Farrand, S. K., Bendich, A. J., Gordon, M. P., and Nester, E. W. (1974). *Agrobacterium tumefaciens* DNA and PS8 bacteriophage DNA not detected in crown gall tumors. *Proc. Natl. Acad. Sci. U.S.A.* 71, 3672–3676. doi: 10.1073/pnas.71.9.3672

Cianfanelli, F. R., Monlezun, L., and Coulthurst, S. J. (2016). Aim, load, and fire: the Type VI secretion system, a bacterial nanoweapon. *Trends Microbiol.* 24, 51–62. doi: 10.1016/j.tim.2015.10.005

Clinical and Laboratory Standards Institute (2012). *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria that Grow Aerobically. M07A9*, 9th Edn. Wayne, PA: Clinical and Laboratory Standards Institute.

Collins, A. J., Fullmer, M. S., Gogarten, J. P., and Nyholm, S. V. (2015). Comparative genomics of *Roseobacter* clade bacteria isolated from the accessory nidamental gland of *Euprymna scolopes*. *Front. Microbiol.* 6:123. doi: 10.3389/fmicb.2015.00123

Collins, A. J., LaBarre, B. A., Won, B. S., Shah, M. V., Heng, S., Choudhury, M. H., et al. (2012). Diversity and partitioning of bacterial populations within the accessory nidamental gland of the squid *Euprymna scolopes*. *Appl. Environ. Microbiol.* 78, 4200–4208. doi: 10.1128/AEM.07437-11

Collins, A. J., and Nyholm, S. V. (2011). Draft genome of *Phaeobacter gallaeciensis* ANG1, a dominant member of the accessory nidamental gland of *Euprymna scolopes*. *J. Bacteriol.* 193, 3397–3398. doi: 10.1128/JB.05139-11

Cude, W. N., Mooney, J., Tavanaei, A. A., Hadden, M. K., Frank, A. M., Gulvik, C. A., et al. (2012). Production of the antimicrobial secondary metabolite indigoidine contributes to competitive surface colonization by the marine *Roseobacter Phaeobacter* sp. strain Y4I. *Appl. Environ. Microbiol.* 78, 4771–4780. doi: 10.1128/AEM.00297-12

Cude, W. N., Prevatte, C. W., Hadden, M. K., May, A. L., Smith, R. T., Swain, C. L., et al. (2015). Phaeobacter sp. strain Y4I utilizes two separate cell-to-cell communication systems to regulate production of the antimicrobial indigoidine. *Appl. Environ. Microbiol.* 81, 1417–1425. doi: 10.1128/AEM.02551-14

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 5:e11147. doi: 10.1371/journal.pone.0011147

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772. doi: 10.1038/nmeth.2109

Durand, E., Cambillau, C., Cascales, E., and Journet, L. (2014). VgrG, Tae, Tle, and beyond: the versatile arsenal of Type VI secretion effectors. *Trends Microbiol.* 22, 498–507. doi: 10.1016/j.tim.2014.06.004

Edenborn, H. M., and Litchfield, C. D. (1985). Glycolate metabolism by *Pseudomonas* sp., *strain S227*, isolated from a coastal marine sediment. *Mar. Bio.* 88, 199–205. doi: 10.1007/BF00397167

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Flórez, L. V., Biedermann, P. H., Engl, T., and Kaltenpoth, M. (2015). Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. *Nat. Prod. Rep.* 32, 904–936. doi: 10.1039/c5np00010f

Furrow, F. B., Amsler, C. D., McClintock, J. B., and Baker, B. J. (2003). Surface sequestration of chemical feeding deterrents in the Antarctic sponge Latrunculia apicalis as an optimal defense against sea star spongivory. *Mar. Bio.* 143, 443–449. doi: 10.1007/s00227-003-1109-5

Geng, H., Bruhn, J. B., Nielsen, K. F., Gram, L., and Belas, R. (2008). Genetic dissection of tropodithietic acid biosynthesis by marine roseobacters. *Appl. Environ. Microbiol.* 74, 1535–1545. doi: 10.1128/AEM.02339-07

Grigioni, S., Boucher-Rodoni, R., Demarta, A., Tonolla, M., and Peduzzi, R. (2000). Phylogenetic characterisation of bacterial symbionts in the accessory nidamental glands of the sepioid *Sepia officinalis* (Cephalopoda: Decapoda). *Mar. Biol.* 136, 217–222. doi: 10.1007/s002270050679

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Hood, R. D., Singh, P., Hsu, F., Güvener, T., Carl, M. A., Trinidad, R. R. S., et al. (2010). A Type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* 7, 25–37. doi: 10.1016/j.chom.2009.12.007

Jiang, F., Waterfield, N. R., Yang, J., Yang, G., and Jin, Q. (2014). A *Pseudomonas aeruginosa* type VI secretion phospholipase D effector targets both prokaryotic and eukaryotic cells. *Cell Host Microbe* 15, 600–610. doi: 10.1016/j.chom.2014.04.010

Kaufman, M., Ikeda, Y., Patton, C., van Dykhuizen, G., and Epel, D. (1998). Bacterial symbionts colonize the accessory nidamental gland of the squid Loligo opalescens via horizontal transmission. *Biol. Bull.* 194, 36–43. doi: 10.2307/1542511

Martens, T., Gram, L., Grossart, H. P., Kessler, D., Müller, R., Simon, M., et al. (2007). Bacteria of the *Roseobacter* clase show potential for secondary metabolite production. *Microb. Ecol.* 54, 31–42. doi: 10.1007/s00248-006-9165-2

McFall-Ngai, M. J. (2014). The importance of microbes in animal development: lessons from the squid-vibrio symbiosis. *Ann. Rev. Microbiol.* 68, 177–194. doi: 10.1146/annurev-micro-091313-103654

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Newton, R. J., Griffin, L. E., Bowles, K. M., Meile, C., Gifford, S., Givens, C. E., et al. (2010). Genome characteristics of a generalist marine bacterial lineage. *ISME J.* 4, 784–798. doi: 10.1038/ismej.2009.150

Nyholm, S. V., and McFall-Ngai, M. J. (1998). Sampling the light-organ microenvironment of *Euprymna scolopes*: description of a population of host cells in association with the bacterial symbiont Vibrio fischeri. *Biol. Bull.* 195, 89–97. doi: 10.2307/1542815

Onaka, H., Mori, Y., Igarashi, Y., and Furumai, T. (2011). Mycolic acid-containing bacteria induce natural-product biosynthesis in *Streptomyces* species. *Appl. Environ. Microbiol.* 77, 400–406. doi: 10.1128/AEM.01337-10

Orf, G. S., and Blankenship, R. E. (2013). Chlorosome antenna complexes from green photosynthetic bacteria. *Photosynth. Res.* 116, 315–331. doi: 10.1007/s11120-013-9869-3

Pichon, D., Gaia, V., Norman, M. D., and Boucher-Rodoni, R. (2005). Phylogenetic diversity of epibiotic bacteria in the accessory nidamental glands of squids (Cephalopoda: Loliginidae and *Idiosepiidae*). *Mar. Biol.* 147, 1323–1332. doi: 10.1007/s00227-005-0014-5

Pukatzki, S., Ma, A. T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W. C., et al. (2006). Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1528–1533. doi: 10.1073/pnas.0510322103

Rao, D., Webb, J. S., Holmström, C., Case, R., Low, A., Steinberg, P., et al. (2007). Low densities of epiphytic bacteria from the marine alga Ulva australis inhibit settlement of fouling organisms. *Appl. Environ. Microbiol.* 73, 7844–7852. doi: 10.1128/AEM.01543-07

Ravn, L., Christensen, A. B., Molin, S., Givskov, M., and Gram, L. (2001). Methods for detecting acylated homoserine lactones produced by Gram-negative bacteria and their application in studies of AHL-production kinetics. *J. Microbiol. Methods* 44, 239–251. doi: 10.1016/S0167-7012(01)00217-2

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Ronquist, F., Teslenko, M., Mark, P., van der Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Ruiz, B., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Sánchez, M., et al. (2010). Production of microbial secondary metabolites: regulation by the carbon source. *Crit. Rev. Microbiol.* 36, 146–167. doi: 10.3109/10408410903489576

Sana, T. G., Hachani, A., Bucior, I., Soscia, C., Garvis, S., Termine, E., et al. (2012). The second type VI secretion system of *Pseudomonas aeruginosa* strain PAO1 is regulated by quorum sensing and fur and modulates internalization in epithelial cells. *J. Biol. Chem.* 287, 27095–27105. doi: 10.1074/jbc.M112.376368

Sanchez, L. M., Curtis, M. E., Bracamonte, B. E., Kurita, K. L., Navarro, G., Sparkman, O. D., et al. (2011). Versatile method for the detection of covalently bound substrates on solid supports by DART mass spectrometry. *Org. Lett.* 13, 3770–3773. doi: 10.1021/ol201404v

Schleicher, T., and Nyholm, S. (2011). Characterizing the host and symbiont proteomes in the association between the Bobtail squid, *Euprymna scolopes*, and the bacterium, *Vibrio fischeri*. *PLoS ONE* 6:e25649. doi: 10.1371/journal.pone.0025649

Schmidt, E. W., and Donia, M. S. (2010). Life in cellulose houses: symbiotic bacterial biosynthesis of ascidian drugs and drug leads. *Curr. Opin. Biotechnol.* 21, 827–833. doi: 10.1016/j.copbio.2010.10.006

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089

Schwarz, S., Hood, R. D., and Mougous, J. D. (2010). What is type VI secretion doing in all those bugs? *Trends Microbiol.* 18, 531–537. doi: 10.1016/j.tim.2010.09.001

Seyedsayamdost, M. R., Case, R. J., Kolter, R., and Clardy, J. (2011). The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nat. Chem.* 3, 331–335. doi: 10.1038/nchem.1002

Tong, Y. Y., and Lighthart, B. (1997). Solar radiation is shown to select for pigmented bacteria in the ambient outdoor atmosphere. *Photochem. Photobiol.* 65, 103–106. doi: 10.1111/j.1751-1097.1997.tb01884.x

Tritt, A., Eisen, J. A., Facciotti, M. T., and Darling, A. E. (2012). An integrated pipeline for de novo assembly of microbial genomes. *PLoS ONE* 7:e42304. doi: 10.1371/journal.pone.0042304

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 40, 1–12. doi: 10.1093/nar/gks596

Unterweger, D., Miyata, S. T., Bachmann, V., Brooks, T. M., Mullins, T., Kostiuk, B., et al. (2014). The *Vibrio cholerae* type VI secretion system employs diverse effector modules for intraspecific competition. *Nat. Comm.* 5, 3549. doi: 10.1038/ncomms4549

Van Dyck, S., Flammang, P., Meriaux, C., Bonnel, D., Salzet, M., Fournier, I., et al. (2010). Localization of secondary metabolites in marine invertebrates: contribution of MALDI MSI for the study of saponins in Cuvierian tubules of H. *forskali*. *PLoS ONE* 5:e13923. doi: 10.1371/journal.pone.0013923

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., et al. (2015). antiSMASH 3.0 - a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43, W237–W243. doi: 10.1093/nar/gkv437

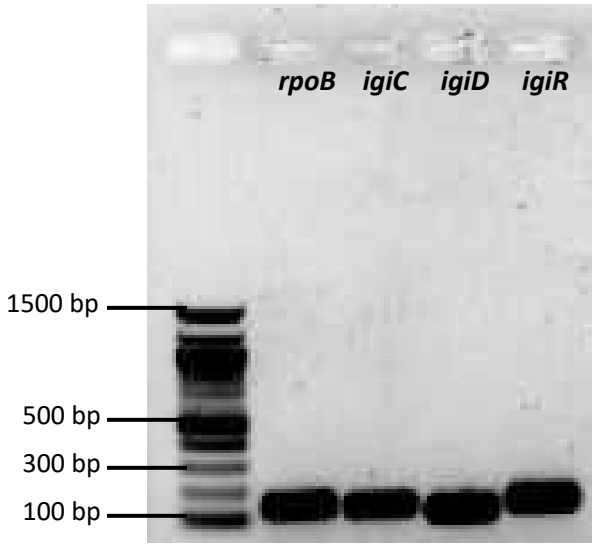
Whistler, C. A., and Ruby, E. G. (2003). GacA regulates symbiotic colonization traits of *Vibrio fischeri* and facilitates a beneficial association with an animal host. *J. Bacteriol.* 185, 7202–7212. doi: 10.1128/JB.185.24.7202

Yu, D., Xu, F., Valiente, J., Wang, S., and Zhan, J. (2013). An indigoidine biosynthetic gene cluster from *Streptomyces chromofuscus* ATCC 49982 contains an unusual IndB homologue. *J. Ind. Microbiol. Biotechnol.* 40, 159–168. doi: 10.1007/s10295-012-1207-9

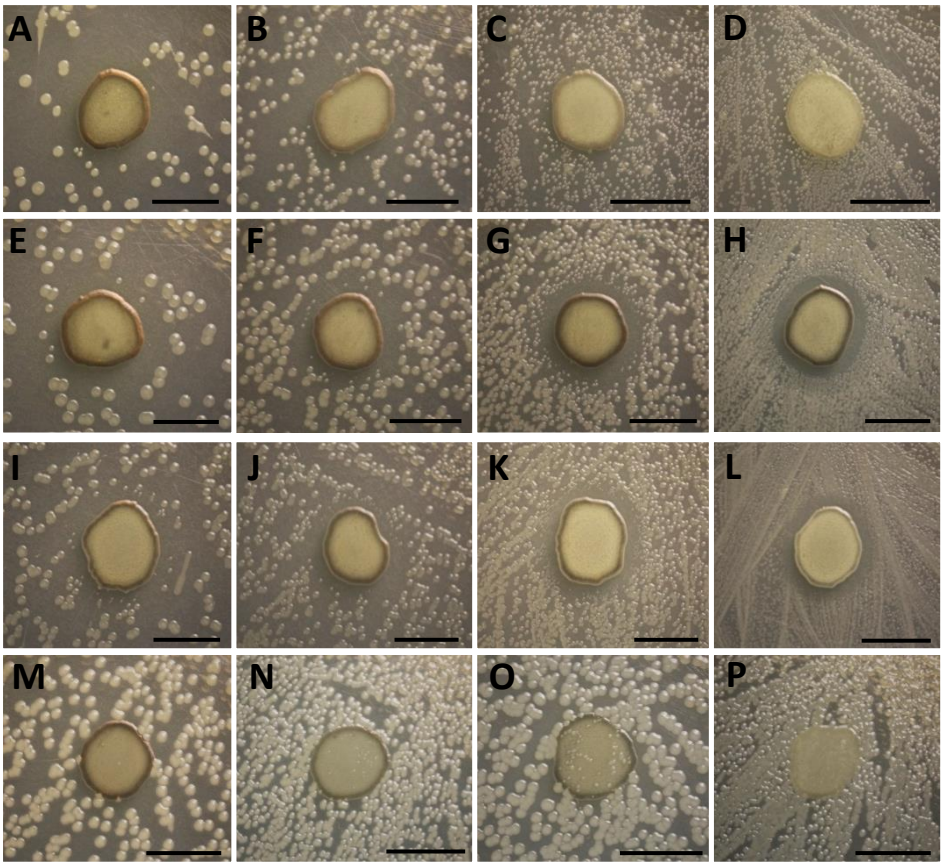
Zan, J., Cicirelli, E. M., Mohamed, N. M., Sibhatu, H., Kroll, S., Choi, O., et al. (2012). A complex LuxR-LuxI type quorum sensing network in a roseobacterial marine sponge symbiont activates flagellar motility and inhibits biofilm formation. *Mol. Microbiol.* 85, 916–933. doi: 10.1111/j.1365-2958.2012.08149.x

Zgoda, J. R., and Porter, J. R. (2001). A convenient microdilution method for screening natural products against bacteria and fungi. *Pharm. Biol.* 39, 221–225. doi: 10.1076/phbi.39.3.221.5934

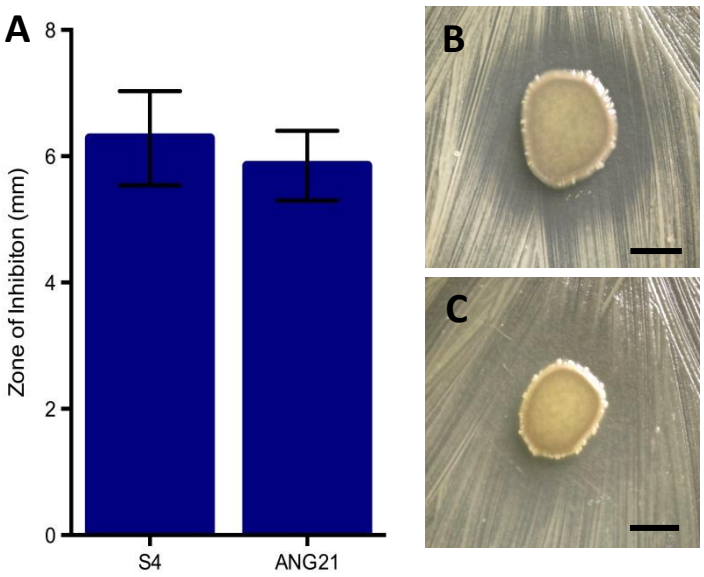
Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



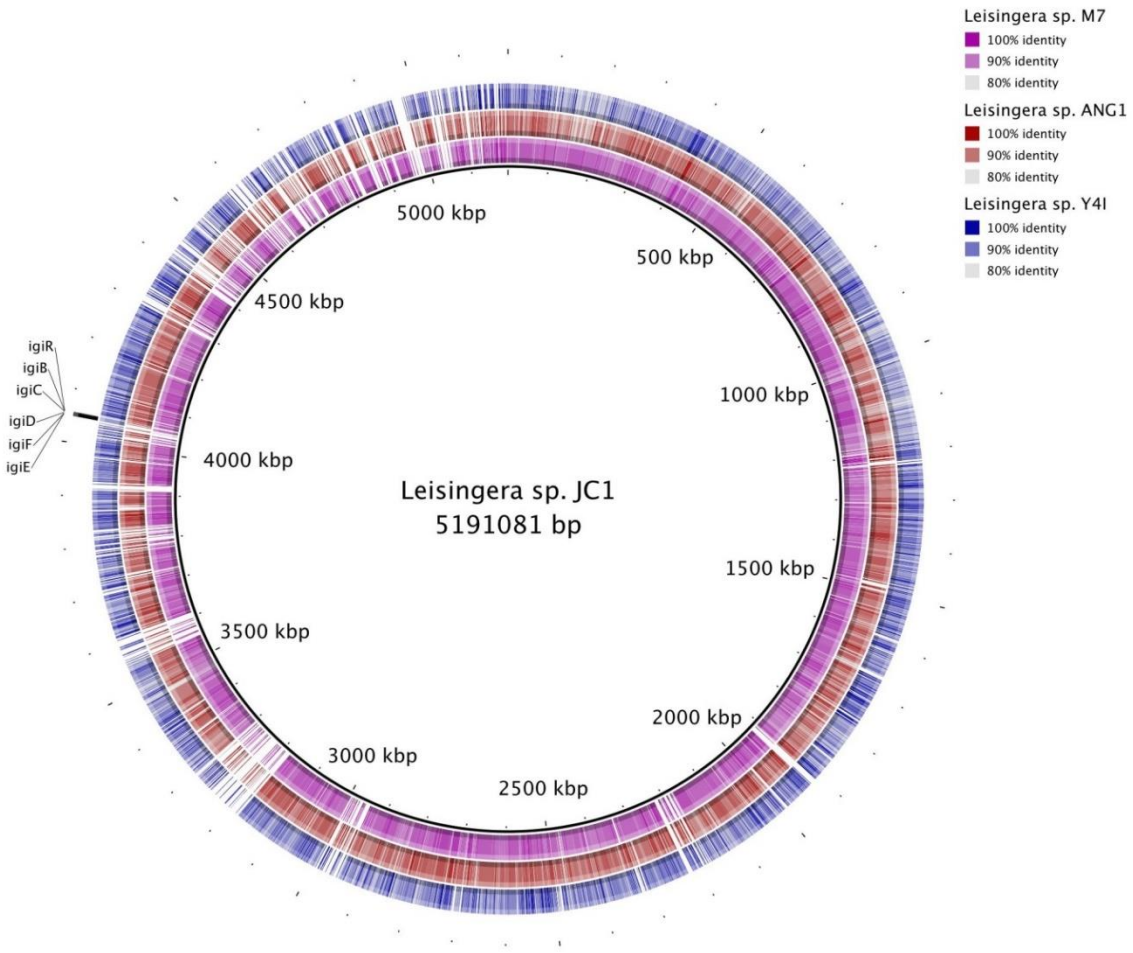
Supplemental Figure 1. Indigoidine biosynthesis genes *igiCDR* detected in *Leisingera* sp. JC1 DNA. The RNA polymerase β subunit, *rpoB,* was used as a positive control for bacterial DNA. Expected product sizes: *rpoB* = 150 bp; *igiC* = 140 bp; *igiD* = 120 bp; *igiR* = 155 bp.



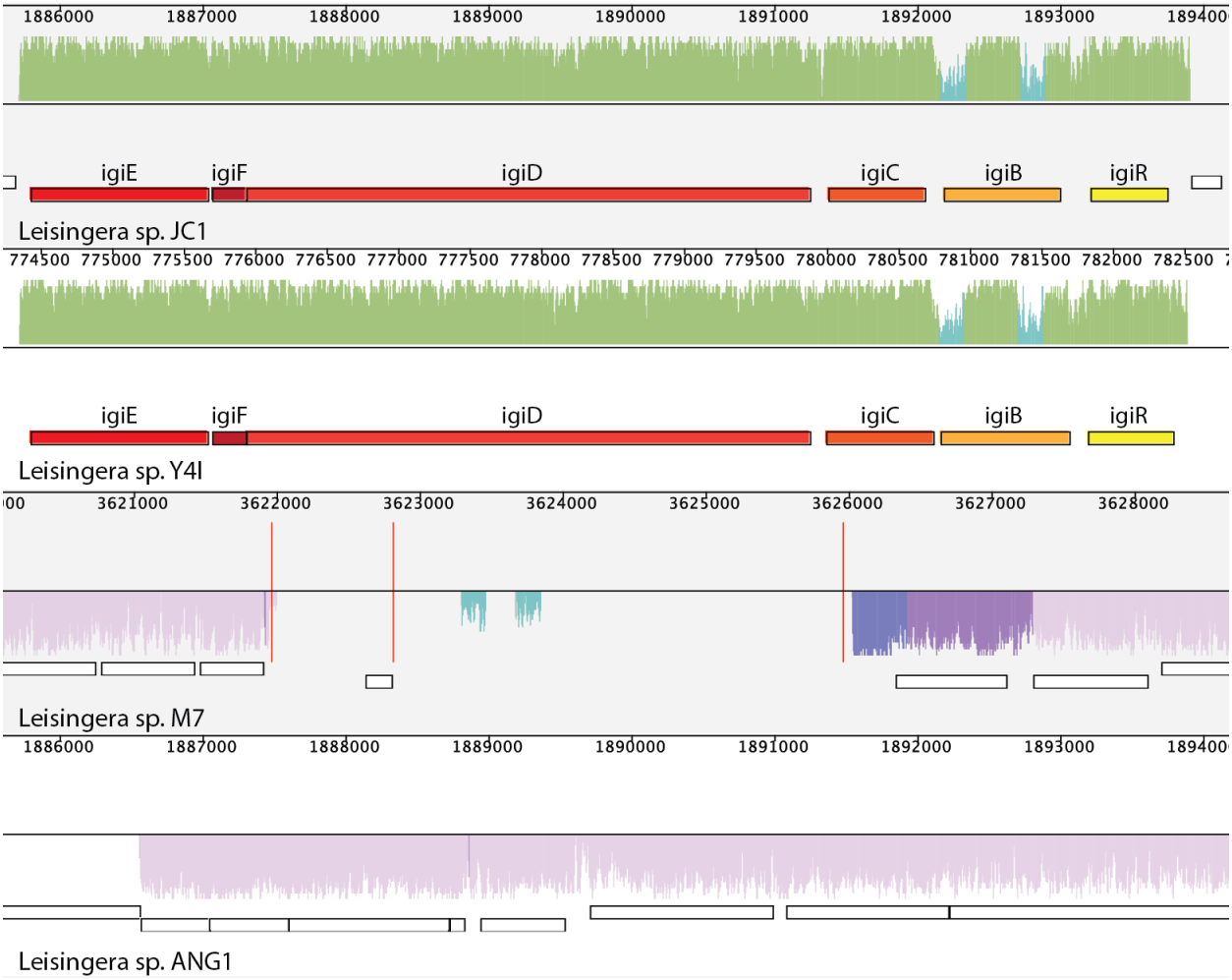
Supplemental Figure 2. ZOI assay images for all dilutions. *Photobacterium leiognathi* at (A) 10^4 CFU/ml, (B) 10^5 CFU/ml, (C) 10^6 CFU/ml, and (D) 10^7 CFU/ml. *Vibrio anguillarum* at (E) 10^4 CFU/ml, (F) 10^5 CFU/ml, (G) 10^6 CFU/ml, and (H) 10^7 CFU/ml. *Vibrio fischeri* at (I) 10^4 CFU/ml, (J) 10^5 CFU/ml, (K) 10^6 CFU/ml, and (L) 10^7 CFU/ml. *Vibrio harveyi* at (M) 10^5 CFU/ml and (N) 10^6 CFU/ml. *Vibrio parahaemolyticus* at (O) 10^5 CFU/ml and (P) 10^6 CFU/ml.



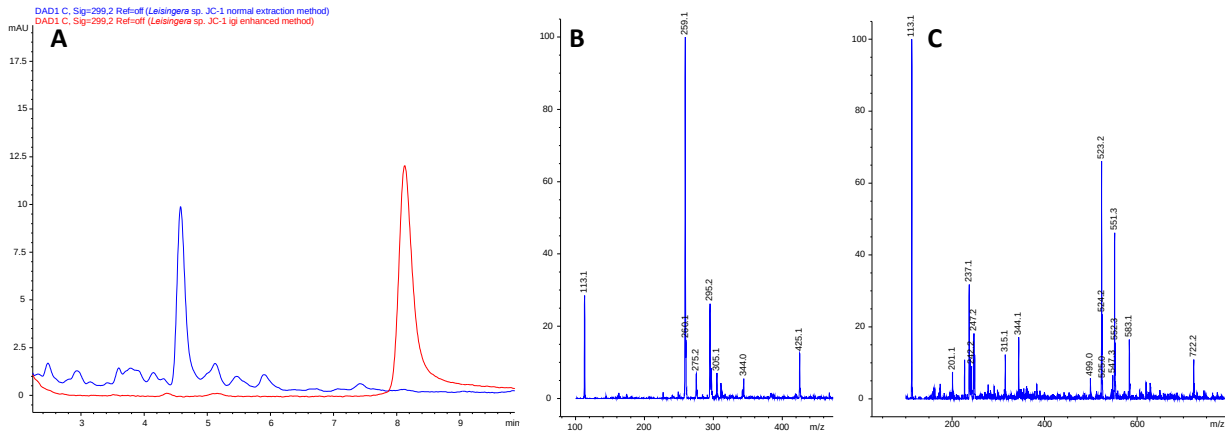
Supplemental Figure 3. *Leisingera* sp. JC1 inhibits ANG isolates *Ruegeria* sp. ANG-S4 and *Muricauda* sp. ANG21. (A) JC1 had an average zone of inhibition (ZOI) of 6.3 mm (±0.7) against *Ruegeria* sp. S4, and an average ZOI of 5.9 mm (±0.6) against *Muricauda* sp. ANG21. Representative images, taken 4 days after plating, of (B) *Ruegeria* sp. ANG-S4 and (C) *Muricauda* sp. ANG21. Scale bars, 5 mm.



Supplemental Figure 4. Genomic comparison of *Leisingera* sp. JC1 with select *Leisingera* species. BLAST comparison of draft genomes of *Leisingera* sp. JC1 with *Leisingera* sp. ANG-M7 (purple inner ring), *Leisingera* sp. ANG1 (red middle ring), and *Leisingera* sp. Y4I (blue outer ring). Location of the indigoidine operon (*igiRBCDFE*) is called out.



Supplemental Figure 5. Genomic comparison of the *Leisingera* sp. JC1 indigoidine operon. Comparison of JC1 indigoidine operon to the operon in *Leisingera* sp. Y4I (top), and to the genomic region in *Leisingera* sp. M7 (middle) and *Leisingera* sp. ANG1 (bottom). The JC1 operon hits to the Y4I operon (as also shown in Table 2). There are no hits when compared to M7 and ANG1, indicating the indigoidine operon is not present in these genomes.



Supplemental Figure 6. Mass spectral confirmation that the normal extraction method contains minimal quantities of indigoidine. LC-MS data was collected on an Agilent system as described in Materials and Methods using an isocratic method of 10% acetonitrile with 0.1% formic acid over 25 minutes at a flow rate of 1 mL/min. Indigoidine eluted at retention time (t_R) 8.2 min as determined by LC-MS. (A) Normal extraction method (blue) was overlaid with indigoidine (igi) enhanced extraction method (red) focusing on UV absorbance 299 nm. The peak eluting at t_R 8.2 min (igi enhanced extraction method) was confirmed to be indigoidine, which was not found in the normal extract. (B) Negative ionization mass spectrum of peak at t_R 4.8 min for normal extraction method, indicating that this peak is not displaced indigoidine since no $[M - H]^-$ of 247 was present. (C) Negative ionization mass spectrum of peak at t_R 8.2 min for normal extraction method. Although there was an ion corresponding to $[M - H]^-$ of 247.2, this was not the most dominant ion and the more quantitative UV analysis in A shows no peak, indicative of little or no indigoidine in the normal extract.

Supplemental Table 1. Primers designed for the detection of indigoidine biosynthesis genes in the *Leisingera* sp. JC1 genome.

Gene	Primer Sequence
igiC Forward	5’ - CGGCAGCAGCGGCAAAATCC - 3’
igiC Reverse	5’ – CGGACGGACGCCGTCAAACA – 3’
igiD Forward	5’ – GGACGGGTCGCGGAATAGCG – 3’
igiD Reverse	5’ – CAGCTGGAGCAGGCAGGCAG – 3’
igiR Forward	5’ – ACCCGCTTTCCAACCGCTCG – 3’
igiR Reverse	5’ – CCTCCGCGCACTGTTGCTGT – 3’

Supplemental Table 2. Whole genome comparisons. Average nucleotide identities (ANI) are presented in the upper triangle and *in silico* DNA-DNA hybridization (*is*DDH) estimates are presented in the lower triangle. ANI and *is*DDH values above the usual cutoffs for species distinction are colored in maroon (ANI, 96%) and red (*is*DDH, 70%).

= ANI ≥ 96% = isDDH ≥ 70%	Leisingera aquimarina DSM 24565	Leisingera methylohalidivorans DSM 14336	Leisingera sp M1	Leisingera sp Vp	Leisingera arcticus DSM 23566	Leisingera caeruleus DSM 24564	Leisingera gallaeciensis DSM 26640	Leisingera inhibens DSM 16374	Leisingera daeponensis DSM 23529	Leisingera sp. Y4I	Leisingera sp. JC1	Leisingera sp. M7	Leisingera sp. ANG1	Leisingera sp. M6	Leisingera sp. S	Leisingera sp. S3	Leisingera sp. S5	Leisingera sp. DT
Leisingera aquimarina DSM 24565		88.4	85.9	86.0	83.4	86.5	84.5	84.3	86.3	86.3	86.3	86.0	86.2	86.2	86.2	86.2	86.3	86.2
Leisingera methylohalidivorans DSM 14336	34.1		85.8	85.8	83.4	86.5	84.3	84.2	86.3	86.3	86.1	85.9	86.0	86.0	86.0	86.0	86.0	86.0
Leisingera sp. M1	26.9	26.7		89.6	83.3	86.4	84.6	84.4	86.3	86.2	86.3	86.2	86.3	86.4	86.3	86.3	86.3	86.3
Leisingera sp. Vp	27.1	26.8	35.7		83.3	86.4	84.6	84.5	86.3	86.3	86.4	86.3	86.4	86.4	86.4	86.3	86.3	86.4
Leisingera arcticus DSM 23566	20.7	20.6	20.3	20.4		83.2	85.1	83.7	83.8	83.4	83.4	83.3	83.3	83.2	83.3	83.2	83.2	83.2
Leisingera caeruleus DSM 24564	28.4	28.2	27.8	27.9	20.4		84.4	84.5	90.6	90.6	87.5	87.1	87.4	87.4	87.4	87.4	87.8	87.5
Leisingera gallaeciensis DSM 26640	21.5	21.3	21.4	21.4	21.9	21.5		89.9	21.4	84.2	84.7	84.4	84.7	84.7	84.7	84.7	84.6	84.7
Leisingera inhibens DSM 16374	20.8	21.2	21.0	21.3	20.8	21.5	38.2		21.2	84.4	84.7	84.5	84.6	84.7	84.7	84.6	84.6	84.5
Leisingera daeponensis DSM 23529	27.9	27.9	27.7	27.6	21.0	40.3	84.5	84.4		98.5	87.4	87.1	87.5	87.5	87.5	87.5	87.5	87.5
Leisingera sp. Y4I	28.0	28.0	27.7	27.7	20.8	40.3	21.3	21.3	86.7		87.4	87.2	87.4	87.5	87.4	87.4	87.5	87.5
Leisingera sp. JC1	27.8	27.4	27.3	27.3	20.5	30.2	21.4	21.4	30.1	30.2		91.7	90.5	90.5	90.5	90.5	90.5	90.5
Leisingera sp. M7	27.2	27.0	26.9	27.0	20.4	29.4	21.4	21.3	29.4	29.5	44.6		90.1	90.1	90.1	90.1	90.1	90.1
Leisingera sp. ANG1	27.7	27.3	27.4	27.3	20.5	30.1	21.4	21.0	30.1	30.1	38.9	37.8		97.9	99.9	99.9	97.8	97.9
Leisingera sp. M6	27.8	27.4	27.4	27.5	20.4	30.1	21.3	21.0	30.2	30.2	38.9	37.7	81.9		97.9	98.0	97.9	97.9
Leisingera sp. S	27.7	27.3	27.4	27.3	20.5	30.0	21.4	21.0	30.1	30.1	38.9	37.7	100.0	82.0		99.9	97.8	97.9
Leisingera sp. S3	27.7	27.3	27.4	27.3	20.5	30.1	21.4	21.0	30.3	30.2	38.9	37.7	100.0	82.1	100.0		97.8	97.9
Leisingera sp. S5	27.9	27.4	27.2	27.3	20.4	30.8	21.5	21.0	30.1	30.2	38.9	37.8	80.8	81.4	80.8	80.8		97.9
Leisingera sp. DT	27.7	27.4	27.3	27.5	20.4	30.1	21.6	20.9	30.2	30.2	38.8	37.7	81.6	81.3	81.7	81.5	81.1	

Supplemental Table 3. Bacterial strains used in this study.

Strain	Characteristics	Source
Leisingera sp. JC1	Isolate from *E. scolopes* egg jelly coats	This study
Agrobacterium tumefaciens NTL4 pZLR4	Reporter strain for detection of HSLs	Singh & Greenstein (2006) (Courtesy of Dr. Stephen Farrand, University of Illinois)
Leisingera sp. ANG1	Isolate from *E. scolopes* ANG	Collins & Nyholm (2011)
Ruegeria sp. ANG-S4	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Ruegeria sp. ANG-R	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-Vp	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-M1	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-DT	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-S	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-S3	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-M6	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-S5	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-M7	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Tateyamaria sp. ANG-S1	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Muricauda sp. ANG21	Isolate from *E. scolopes* ANG	This study
Vibrio fischeri ES114	Isolate from *E. scolopes* light organ	Boettcher & Ruby (1990)
Vibrio harveyi B392	Seawater isolate	Reichelt & Baumann (1973)
Vibrio anguillarum 775		Courtesy of Dr. Joerg Graf, University of Connecticut
Vibrio parahaemolyticus KNH1	Seawater isolate	Nyholm et al. (2000)
Photobacterium leiognathi KNH6	Seawater isolate	Stabb & Ruby (2002)

Supplemental Table 4. Detailed antiSMASH results for *Leisingera* sp. JC1.

Cluster	Type	From – To (bp)	Top homologous gene clusters
1	Homoserine lactone	(Scaffold 1.1) 332,333 – 352,974	43% sim to *Leisingera* sp. ANG1 39% sim to *Leisingera* sp. ANG-VP 41% sim to *Leisingera* sp. ANG-M7
2	Bacteriocin	(Scaffold 14.1) 57,539 – 68,393	29% sim to *Leisingera* sp. ANG-M7 29% sim to *Leisingera* sp. ANG-DT 29% sim to *Leisingera* sp. ANG-M6
3	Other KS	(Scaffold 17.1) 19,628 – 60,623	12% sim to *Citreicella* sp. SE45 12% sim to *Shinella* sp. DD12 SHLA 2c 12% sim to *Brachybacterium paraconglomeratum*
4	T1 PKS	(Scaffold 24.1) 1 - 24,550	45% sim to *Leisingera* sp. ANG-M7 45% sim to *Leisingera* sp. ANG-S3 45% sim to *Leisingera* sp. ANG1
5	Other	(Scaffold 34.1) 13,994 – 41,877	17% sim to *Leisingera* sp. Y4I 14% sim to *Vogesella indigofera* 11% sim to *Gordonia soli* NBRC 108243
6	Siderophore	(Scaffold 63.1) 2,810 – 14,229	18% sim to *Mesorhizobium* sp. L48C026A00 18% sim to *Rhizobium trpici* CIAT 899 plasmid pRtrCIAT899c 16% sim to *Acidovorax radices* N35
7	Other	(Scaffold 79.1) 1 – 9,934	16% sim to *Leisingera* sp. ANG-M7 16% sim to *Leisingera* sp. ANG-M6 16% sim to *Leisingera* sp. ANG1
8	Siderophore	(Scaffold 99.1) 1 – 3,224	7% sim to *Leisingera* sp. ANG1 7% sim to *Leisingera* sp. ANG-S 7% sim to *Leisingera* sp. ANG-S3
9	Siderophore	(Scaffold 114.1) 1 – 1,268	None found

Chapter 2 – Part 2

Accessory nidamental gland and egg jelly coat isolates inhibit marine bacteria *in vitro*: Antibacterial activity, chemical, and genomic analysis of a Hawaiian bobtail squid egg isolate, *Pseudoalteromonas* sp. JC28

Significant Contributions

Significant contributions to this chapter were made by other researchers. Karen Tan and Marcy Balunas (Division of Medicinal Chemistry, Department of Pharmaceutical Sciences, University of Connecticut) generated organic extracts of *Pseudoalteromonas* sp. JC28, performed extract testing, characterized the structures of the bromoalterochromides, and wrote portions of the chapter. Allison Kerwin (Department of Molecular and Cell Biology, University of Connecticut) assisted with zone of inhibition assays and statistics in figure 1. Lydia Abini-Agbomson, Jessica Bertenshaw, and Jaydeen Sewell (University of Connecticut) assisted with initial zone of inhibition testing in table 1.

Abstract

The symbiotic community in the accessory nidamental gland (ANG) of the female Hawaiian bobtail squid, *Euprymna scolopes*, has been postulated to protect the host eggs against pathogens and fouling agents. One member of this community, *Pseudoalteromonas* sp. JC28, was isolated from the egg jelly coat (JC), and was found to produce known bromoalterochromides (BAC's), BAC-A/A' (**1a**/**1b**), and new derivatives, BAC-D/D' (**2a**/**2b**). Extensive NMR analyses allowed complete structural elucidation of **2a**/**2b**, and the absolute stereochemistry was unambiguously confirmed by an optimized Marfey's method to corroborate with the BAC gene cluster, which we identified from the JC28 draft genome. In this study, the potential ecological role of BAC's was investigated by performing *in vitro* antibacterial, nitric oxide (NO) inhibition, antifungal, and cytotoxicity assays. Compounds **1a**/**1b** were active against marine bacteria, *Bacillus algicola* and *Vibrio fischeri*, with MIC's of 7.4 μM and 59.3 μM, respectively. They also inhibited NO production (IC_{50} 16.9 μM) during lipopolysaccharide (LPS)-induced inflammation in RAW264.7 macrophage cells. These results indicate that BACs may provide the producer organism an ecological niche by facilitating host colonization and eliminating competition. Furthermore, the compounds inhibited hyphal growth of the pathogenic fungus, *Fusarium keratoplasticum*, and were cytotoxic against MCF-7 breast cancer cells, suggesting that these polyketide-nonribosomal peptides are promising drug leads, and may also be used for squid egg defense against pathogens and predators. These findings provide a molecular insight into the exquisite mutualistic relationship between the squid host and one of its JC symbionts through bacterial secondary metabolite production.

Introduction

The role symbiotic bacteria play in the reproductive success of their hosts is a field of growing interest. Organisms which deposit eggs without parental care leave their eggs vulnerable to threats of predation, infection, and fouling. To compensate, some insects utilize the compounds produced by bacteria to defend against microbial threats, such as beewolf wasps (47)(46) and houseflies (49). In marine environments, shrimp and lobsters supplement parental care with antifungal compounds produced by bacteria that coat their eggs (54)(55). In the Hawaiian bobtail squid, *Euprymna scolopes*, symbiotic bacteria are deposited into the jelly coat surrounding developing embryos to serve as a defense from microbial pathogens during the month long external embryogenesis of *E. scolopes* eggs (Appendix I).

Female bobtail squid house a diverse bacterial consortium in a reproductive gland known as the accessory nidamental gland (ANG) (69). This gland is attached to the nidamental gland, which produce a jelly that surrounds the yolk sac of fertilized eggs. The ANG deposits bacteria into this jelly coat (JC), which are hypothesized to produce bioactive chemicals that can defend the eggs from bacterial and fungal infections or competitively exclude pathogens from settling on the egg surface. In *E. scolopes*, the ANG and JC communities are composed of *Alphaproteobacteria* (66% ANG/71% JC), *Verrucomicrobia* (29%/8%), *Gammaproteobacteria* (4%/4%), *Flavobacteriia* (0.9%/12%) (72). Genomic analyses of some of these *Alphaproteobacteria* isolate genomes have revealed the potential to make several secondary metabolites, such as threw polyketide synthase (PKS) or nonribosomal peptide synthetases (NRPS) pathways (103)(85). Previous research has shown that one JC isolate, *Leisingera* sp. JC1, can produce the antimicrobial pigment, indigoidine, and inhibit pathogenic vibrios *in vitro*

(85). Aside from indigoidine, however, there have been no other described bioactive compounds isolated from ANG/JC bacteria.

To further characterize the inhibitory activity of the ANG/JC bacterial isolates in culture, an *in vitro* screen was performed on 19 strains. As a result, one isolate, *Pseduoalteromonas* sp. JC28, was found to have strong inhibitory activity against several Gram positive and Gram negative marine bacteria. This isolate and its organic extract were previously shown to inhibit the fungal pathogen, *Fusarium keratoplasticum* (Appendix I). Due to this broad antimicrobial activity, *Pseudoalteromonas* sp. JC28 was chosen for genomic and chemical analyses.

Purification of the JC28 bacterial culture led to the isolation of the major compounds, the bromoalterochromides (BACs). These brominated derivatives of alterochromides were first identified by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) from the yellow-pigmented, sponge tissue-derived *Pseudoalteromonas maricaloris* KMM636 (111), but an attempt to the full spectroscopic characterization of their structures was not published until two years later (112), probably owing to low yields and compound instability. Relying on mass spectrometric methods, they have also been identified from several *Pseudoalteromonas* strains (113)(114)(115), indicating widespread metabolite expression within this bacterial genus. However, the complete structural elucidation of their analogues have been neglected, and inconsistencies in terms of their chemically-determined and bioinformatically-predicted absolute stereochemistry have been observed (112)(116)(117). Therefore, one of the aims of this study is to completely characterize the structures of the isolated BACs, and to make necessary corrections to their reported absolute stereochemistry by applying high-resolution nuclear magnetic resonance (NMR) spectroscopic and C3 Marfey techniques (118). Bioactivity-profiling of

Pseudoalteromonas sp. JC28's BACs may also provide insights into how bacterial secondary metabolites contribute to the *E. scolopes* ANG/JC symbiosis.

Materials and Methods

Bacterial strains, identification, and culture conditions

Seven of the strains used in this study previously isolated from the ANGs of female *E. scolopes* (103)(85) (Table S1). Eight more strains were isolated from ANGs of 3 females for this study as previously described (103) (Table S1). Four strains were isolated from egg jelly coats as previously described (85) (Table S1). Strains isolated in this study were identified to the genus level by Sanger sequencing of the 16S rRNA gene. A colony PCR was performed using GoTaq Master Mix Green (Promega, Madison, WI) with 27F and 1492R universal primers (119). The PCR cycle was 95˚C for 3 min, 30 cycles (95˚C 30 sec, 55˚C 30 sec, 72˚C 1 min 30 sec), final elongation of 72˚C for 10 min. The PCR product was purified using ExoSap It (Affymetrix, Santa Clara, CA) and prepared for sequencing using Big Dye Terminator v.1.1 (Applied Biosystems, Foster City, CA). Sanger sequencing was performed by the Biotech Center at the University of Connecticut. All strains were grown on seawater tryptone (SWT) media (5 g/L tryptone, 3 g/L yeast extract, 3 mL/L glycerol, 700 mL/L artificial seawater [Instant Ocean sea salts], 15 g/L agar, 300 mL/L DI water).

Zone of Inhibition (ZOI) assay

The ability of ANG/JC isolates to inhibit Gram positive (*Bacillus megaterium* CNJ 778 and *Exiguobacterium aesteurii* CNJ 771) and Gram negative (*Vibrio fischeri* ES114, *Vibrio anguillarum* 775, *Vibrio parahaemolyticus* KNH1, *Vibrio harveyi* B392, and *Photobacterium*

leiognathi KNH6) marine bacteria was tested *in vitro* on SWT agar plates. In an initial screen, a stationary phase culture of each marine target strain was grown up in SWT broth and a lawn was spread on the surface of an agar SWT plate at a concentration of ~10^8 CFU/ml. These lawns dried for 15 minutes and then a stationary phase culture of the ANG/JC isolate (~10^8 CFU/ml) was spotted (10 µl) onto the center of the plate. Plates were incubated at 28˚C for 24 hours. If a zone of clearance was observed around the ANG/JC isolate, this was marked as positive for inhibition. If no clearance zone was observed, the strain was marked as negative for inhibition. Each isolate was tested against each marine target strain in triplicate. Zone of inhibition is defined as a discrete area (of at least 15 mm^2, after colony area is subtracted) around a colony where there was no observable growth of the lawn strain.

After the initial ZOI screen, the JC isolate, *Pseudoalteromonas* sp. JC28, was chosen for further testing based on its ability to inhibit all marine target strains. To see if there was a density-dependent effect of this inhibitory activity, JC28 was tested against the same marine strains at three lower lawn densities, 10^4, 10^5, and 10^6 CFU/ml. All strains were grown as described above. JC28 was grown to stationary phase (10^8 CFU/ml) and spotted (10 µl) on the center of each lawn. Plates were incubated for 24 hours at 28˚C and imaged. Inhibition was quantified by measuring the area (cm^2) of each zone of inhibition and subtracting out the area of the JC28 colony using the image analysis program, FIJI (120). Each test was performed in experimental and technical triplicates and the ZOI area of each technical replicate was averaged. For each target lawn strain, statistical significance between ZOI at each lawn density was determined using one-way ANOVAs. If the results of this analysis indicated significant differences, a multiple comparisons post hoc Tukey test was performed to determine which lawn densities were significantly different.

Genome sequencing

Pseudoalteromonas sp. JC28 was grown in 3.0 ml of SWT broth, 200 rpm, at 30˚C for 7 hours. Genomic DNA was extracted from 5 x 10^8 CFU/ml of culture using the MasterPure DNA Purification kit (Epicentre, Madison, WI) following the manufacturer's protocol. DNA quantity was measured with a high sensitivity DNA Qubit fluorimeter (Invitrogen, Carlsbad, CA) and quality was checked by absorbance measured on a Nanodrop spectrophotometer (Thermo Scientific, Waltham, MA) and 1% agarose gel. DNA was prepped for sequencing using the Nextera XT DNA library prep kit (Illumina, San Diego, CA). Library yield was checked with Qubit and size and quality measured on a high sensitivity DNA bioanalyzer chip (Agilent Technologies, Santa Clara, CA). Sequencing was performed on an Illumina MiSeq, reagent kit v.2 with 500 cycles at the Microbial Analysis, Resources, and Services (MARS) facility at the University of Connecticut.

Sequencing yielded 3,976,074 raw, paired end reads (785.79 MB), which were uploaded to the U.S. Department of Energy Kbase server version 1.5.2 (https://kbase.us/) for analysis. Raw reads were trimmed and adaptors removed using Trimmomatic (121). Reads were assembled using the A5 assembler (122). This assembly was uploaded to the Rapid Annotation using Subsystem Technology (RAST) server (123) for annotations and to the antiSMASH server v.4.0 (124) for secondary metabolite biosynthetic cluster prediction. The draft genome assembly has been deposited in DDBJ/EMBL/GenBank under accession QOKX00000000 and is described in BioProject PRJNA480278. The version described in this paper is version QOKX01000000.

Multi-locus sequence analysis (MLSA)

To determine the phylogenetic placement of *Pseudoalteromonas* sp. JC28 within the pseudoalteromonads, a MLSA tree was assembled using 34 published *Pseudoalteromonas* genomes. Six single-copy housekeeping genes were selected for multi-locus sequence analysis based on a previous study (125), including ATP synthase subunit beta, *atpD*, DNA gyrase subunit B, *gyrB*, rod shape-determining protein, *mreB*, RNA recombinase alpha subunit, *recA*, RNA polymerase sigma factor, *rpoD*, and DNA topoisomerase 1, *topA*. Genomes were downloaded from Genbank using a python script (https://github.com/kblin/ncbi-genome-download). Amino acid sequences for each gene were obtained by a tblastn search using the amino acid sequence from *E. coli* strain K12 obtained from Uniprot. Extracted sequences were aligned using MUSCLE (126) with default parameters. The multiple alignments were concatenated in Geneious v.10 (Biomatters Inc., Newark, NJ). The best evolutionary model of substitution was chosen by ModelFinder (127) in the IQtree webserver (128) as GTR + F + Invar + Gamma 4. A maximum likelihood tree was constructed using IQtree v.1.6.5 (129) with 1000 bootstraps (130). The tree was visualized using the Interactive Tree of Life (iTOL) v.4 webserver (https://itol.embl.de/) (131).

Presence of pigmentation in a strain was noted from the following articles and added to the tree: (citations in order of top strain to bottom strain in the tree) (132), (133), (134), (135), (136), (137), (138), (137), this study, (139), (140), (141), (141), (137), (142), (143), (144), (145), (146), (147), (148), (149), (150), (151), (137), (137), (152), (153), (137), (154), (137), (155), (156), (142), (157), (157).

Presence of the bromoalterochromide biosynthetic genes in a strain was determined by a tBLASTn search. The amino acid sequence for each bromoalterochromide (BGC0000314)

biosynthetic gene was obtained from the Minimum Information about a Biosynthetic Gene cluster (MIBiG) server (158). This cluster was identified from *Pseudoalteromonas piscicida* JCM 20779 (D. D. Nguyen et al., 2013) and contains 14 genes (PpisJ2_010100002366 – PpisJ2_010100002431). A strain was marked as possessing the bromoalterochromide cluster if all 14 genes were present in the same order and had a high amino acid similarity (≥87%).

Extraction of *Pseudoalteromonas* sp. JC28 and isolation of bromoalterochromides

A culture of JC28 was prepared by inoculating the bacteria into 30 mL of marine broth (Difco Laboratories, Sparks, MD, USA) and incubating at 28˚C overnight at 200 rpm. A 7.5 mL aliquot of the culture was added into four 1.0 L baffled flasks, each containing 500 mL of marine broth, before incubation at 28˚C for 72 h at 200 rpm, protected from light. Solvents for extraction and semi-preparative HPLC purification were ACS grade and HPLC grade, respectively, from Sigma Aldrich (St. Louis, MO, USA). The extraction procedure was performed away from direct light exposure.

The cultures in the four flasks were combined and centrifuged at 6,300 rpm at 4˚C for 10 min, and the supernatant was separated from the pellet by decantation. The aqueous supernatant was extracted with *n*-hexane (3 x 500 mL), followed by EtOAc (12 x 300 mL). Meanwhile, the pellet was soaked in 100 mL of acetone overnight. The acetone extract was vacuum filtered and evaporated to dryness under reduced pressure before partitioning between *n*-hexane (200 mL) and 90% aqueous MeOH (3 x 200 mL). Bromoalterochromides were detected by LC-MS analysis in the EtOAc extract from the supernatant, as well as the 90% aqueous MeOH extract from the pellet. LC-MS conditions: 150 x 4.6 mm Agilent Eclipse-XDB C18, 5 μm, 0.7 mL/min, mobile phase A = 0.1% HCOOH in water, mobile phase B = 0.1% HCOOH in CH_3CN, column

temperature = 24°C, and either a long gradient run (25% B to 100% B over 75 min) or a short gradient run (40% B to 100% B over 30 min). These two extracts were further purified by isocratic elution with 40% aqueous CH_3CN at 2.0 mL/min using a semi-preparative HPLC column (250 x 9.6 mm Agilent Eclipse XDB-C18 column, 5 μm), and monitoring at 395 nm. The EtOAc extracted yielded 0.8 mg of **1a**/**1b** and 0.4 mg of **2a**/**2b**, while the 90% aqueous MeOH extracted yielded 3.3 mg of **1a**/**1b** and 0.7 mg of **2a**/**2b**. Isolated compounds were stored in amber-colored vials under an inert atmosphere at -20°C.

Characterization and confirmation of bromoalterochromide (BAC) structures

The planar structures of the bromoalterochromides (BACs) were determined by extensive NMR analyses (^{1}H, ^{1}H-^{1}H COSY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, and NOESY) using a Varian INOVA 600 MHz instrument (Agilent Technologies, Santa Clara, CA, USA), equipped with a cryoprobe to enhance sensitivity. The NMR spectra were referenced to the solvent peaks of DMSO-d_6 (δ_H = 2.50 ppm, δ_C = 39.52 ppm). High-resolution mass data were obtained using a Waters XeVo G2-XS QToF mass spectrometer (Waters Corporation, Milford, MA, USA), equipped with a UPLC column (50 x 2.1 mm HSS T3 C18, 1.8 μm), and eluted using a gradient of mixtures of mobile phase A (0.1% aqueous HCOOH) and B (0.1% HCOOH in CH_3CN). Program setting: 5% B over 0.5 min, followed by linear gradient from 5% B to 60% B for 3.5 min, 60% B to 98% B for 4 min, and hold at 98% B for 1 min. The samples were dissolved in MeOH at 1 mg/mL, and 10 μL was injected for analysis.

Determination of BACs absolute configuration

All standards and reagents were obtained from Sigma Aldrich, except for DL-allo-Thr from Wako (Wako Chemicals USA, Inc., Richmond, VA, USA), and D-allo-Thr and 1-fluoro-2,4-dinitrophenyl-5-L- leucinamide (L-FDLA) from TCI (Tokyo Chemical Industry America, Portland, OR, USA). The LC-MS data were collected on an Agilent ESI single quadrupole mass spectrometer coupled to an Agilent 1260 HPLC system with a G1311 quaternary pump, G1322 degasser, and a G1315 diode array detector.

The BACs (200 μg) were heated at 105°C with 200 μL of 6-N HCl in a sealed flask for 3-4 h or until the reaction mixture turned colorless. The mixture was evaporated to dryness under nitrogen, and the resulting residue was dissolved in 50 μL of water and 20 μL of 1M $NaHCO_3$. Then, 40 μL of 1% L-FDLA or 1% 1-fluoro-2,4-dinitrophenyl-5-L-alaninamide (L-FDAA) in acetone was added prior to incubation at 35°C for 1 h. The reaction was quenched by addition of 20 μL of 1N HCl, then diluted with 370 μL of MeOH, from which 25 μL was injected for LC-MS analysis. The alkalinified solutions of the L- and D-forms of standard amino acids (100 μg), Val, Leu, Ile, Thr, Asn, and Asp, were derivatized in the same manner as the natural products. After quenching with 20 μL of 1N HCl, they were diluted with 810 μL of MeOH, from which 25 μL was injected for LC-MS analysis. LC-MS condition A (for FDLA-derived samples): 150 x 4.6 mm Agilent Zorbax-SB C3, 5 μm, 0.7 mL/min, mobile phase A = 5% CH_3CN with 1% HCOOH in water, mobile phase B = 5% CH_3CN with 1% HCOOH in MeOH, 30% B to 100% B over 55 min, column temperature = 40°C. LC-MS condition B (for FDAA-derived samples): 150 x 4.6 mm Agilent Zorbax-SB C3, 5 μm, 0.7 mL/min, mobile phase A = 7.5% CH_3CN with 1% HCOOH in water, mobile phase B = 7.5% CH_3CN with 1% HCOOH in MeOH, 12.5% B to 57.5% B over 80 min, column temperature = 25°C.

BAC antibacterial assay

BACs were tested against the Gram positive bacterium, *Bacillus algicola* CNJ 803, and the Gram negative bacterium, *V. fischeri* ES114 in a 96-well liquid assay. Prior to the assay, *B. algicola* was grown on nutrient agar (BD, Franklin Lakes, NJ), and *V. fischeri* on SWT agar, overnight at 28˚C. Inoculums of the target strains were prepared by adding colonies to nutrient broth or SWT broth and adjusting the OD_{600} to 0.08-0.10. A master mix was then prepared by mixing 7.84 mL of the appropriate broth, 6.40 mL water, and 1.60 mL of the inoculum. Into a 96-well round-bottomed plate (Corning Costar, Corning, NY, USA), 199 µL of the master mix was added with 1.0 µL of the BAC sample. The plate was incubated overnight at 28˚C, and with constant shaking at 200 rpm in the case of *V. fischeri*. The wells were observed visually for presence or absence of bacterial growth to determine the minimum inhibitory concentration (MIC). Final testing concentrations were 250 µg/mL for extracts, and 0.39 to 50 µg/mL for the purified compounds, which were all dissolved in DMSO. Positive controls were 2.5 µg/mL of vancomycin (Sigma Aldrich, St. Louis, MO) for *Bacillus*, and 2.5 µg/mL of chloramphenicol (Acros Organics, Morris, NJ, USA) for *Vibrio*, while the negative control was DMSO with a 0.5% final concentration. The assays were done in technical triplicates and at least two experimental replicates to confirm results.

BAC nitric oxide (NO) inhibition assay

An aliquot (197 µL) of RAW264.7 macrophage cells in DMEM, high glucose, pyruvate (Thermo Fisher Scientific), supplemented with 10% fetal bovine serum (Thermo Fisher Scientific) and 1% penicillin-streptomycin (Sigma Aldrich), was seeded into a 96-well flat-

bottomed plate (Corning Costar) at a density of $2.5x10^4$ cells/well, and incubated at 37°C in an atmosphere of 5% CO_2. After 24 h, 2 μL of LPS (final concentration, 100 ng/mL) and 1.0 μL of BAC sample (dissolved in DMSO) or control was added into the well, and incubation continued under the same conditions for another 24 h. Then, 100 μL of the supernatant was transferred into another flat-bottomed 96-well plate, to which 50 μL of 1% sulfanilamide (Sigma Aldrich) in 5% phosphoric acid (Aqua Solutions, Inc., Deer Park, TX, USA) was added. After 5 min, 50 μL of 0.1% aqueous *N*-(1-naphthyl)-ethylenediamine (NED) was mixed into the wells. The plates were read after another 5 min at 540 nm using a Synergy H1 Hybrid Reader (Biotek, Winooski, VT, USA). NO production was expressed as a percentage of the ratio between the sample absorbance and that of the negative control, DMSO. The positive control was 10 μM of suberoylanilide hydroxamic acid (SAHA) (Sigma Aldrich).

BAC antifungal disc assay

Fusarium keratoplasticum FSSC-2g (Appendix I) was spread onto an SWT agar plate and allowed to grow at 28°C for 24-48 h. An inoculum with OD_{600} of 0.01 was prepared from the colonies on the plate in SWT broth. Sterile, 6 mm paper discs (Difco Laboratories) were placed on top of an SWT agar plate, to which 10 μL of BAC sample or control was added. The discs were allowed to dry for about 5 min before adding the fungal inoculum onto them. The plates were then incubated at 28°C and observed for hyphal growth every 24 h for 3 days. Compounds **1a**/**1b** dissolved in DMSO were tested at 10 μg to 100 μg/disc. The positive control was cycloheximide (Sigma Aldrich) at 68 μg/disc, and the negative control was 10 μL of DMSO (Fisher Scientific, Fair Lawn, NJ, USA).

BAC cytotoxicity assay

MCF-7 breast cancer cells were seeded in DMEM (1X), low glucose, pyruvate (Thermo Fisher Scientific), supplemented with 10% fetal bovine serum (Thermo Fisher Scientific) and 1% penicillin-streptomycin (Sigma Aldrich). Then,199 μL of the cell mixture was transferred into a 96-well flat-bottomed plate (Corning Costar) at a density of $2.5x10^4$ cells/well and incubated at 37°C in an atmosphere of 5% CO_2. After 24 h, 1.0 μL of BAC (dissolved in DMSO) or control was added to the well, and incubated under the same conditions for another 24 h. Then, 38 μL of 2 mg/mL [3,(4,5-dimethylthiazol-2- yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS) (Promega Corporation, Fitchburg, WI, USA) and 2 μL of 0.9 mg/mL phenazine methosulfate (PMS) (Acros Organics) were mixed into the wells, and allowed to react with the cells for 3 h at 37°C. The absorbance was measured at 490 nm using a plate reader. The positive control used was doxorubicin (Fisher Scientific) at 10 μM, and DMSO was used as the negative control. Cell viability was calculated as the percentage of the absorbance ratio between the sample well and DMSO. The assay was done in technical triplicates with at least two experimental replicates to verify reproducibility.

Results

Zone of Inhibition Assays

An initial screen of 19 ANG/JC isolates for *in vitro* antibacterial activity revealed two strains, *Leisingera* sp. ANG59 and *Pseudoalteromonas* sp. JC28, which could inhibit one or more of the marine bacteria tested (**Table 1**). All other ANG/JC strains tested did not form any zones of inhibition when spotted on the surface of marine Gram negative and Gram positive lawns at a high density (10^8 CFU/ml). The *Alphaproteobacteria* strain, *Leisingera* sp. ANG59,

produced moderate zones of inhibition against two of the vibrio strains, *V. fischeri* and *V. anguillarum*. These results are comparable to those seen by a previous *E. scolopes* jelly coat isolate, *Leisingera* sp. JC1, which was previously reported to inhibit select vibrio strains in a similar bioassay (85). The strain ANG59 also produces a dark-blue pigment and contains the indigoidine biosynthetic genes associated with antimicrobial activity in JC1 (85).

The *Gammaproteobacteria* JC isolate, *Pseudoalteromonas* sp. JC28, showed the strongest inhibition of all isolates tested and was able to inhibit all five of the Gram negative target bacteria and both Gram positive target bacteria. Due to this general inhibitory activity, JC28 was selected for further antibacterial testing to determine if greater inhibition was observed at lower target lawn densities. Testing lower lawn densities (10^4-10^6 CFU/ml) would be more representative of the concentrations of bacteria in the seawater which these isolates are exposed to over the course of egg development.

Significantly greater inhibition was seen between 10^4-10^5 CFU/ml and 10^5-10^6 CFU/ml lawn densities for all test strains except the 10^4-10^5 CFU/ml densities of *P. leiognathi* and the 10^5-10^6 CFU/ml densities of *E. aesteuri* (**Figure 1**). The greatest inhibition was seen against *V. anguillarum* (ZOI = 11.1 cm^2; **Table S2**) at the lowest lawn density, followed by the Gram positive, *E. aesteuri* (ZOI = 5.3 cm^2; **Table S2**) at the lowest lawn density.

Genomic and phylogenetic analyses

The genome of *Pseudoalteromonas* sp. JC28 was sequenced on the Illumina MiSeq, assembled into 40 scaffolds using the A5 assembly pipeline (122), and annotated with RAST (123). The genome size is 5.53 Mb with a GC content of 43.2%, N50 of 509,001 bp, and 4,894 coding sequences. In the MLSA analysis, *Pseudoalteromonas* sp. JC28 was placed in a well-

supported clade containing mostly pigmented pseudoaltermonads isolated from seawater samples (**Fig. 2**). The closest strain to JC28 was *Pseudoalteromonas flavipulchra* JG1. This clade also contained five *Pseudoalteromonas piscicida* strains and one *Pseudoalteromonas elyakovii* strain. With a few exceptions, most strains examined fell into either a pigmented or non-pigmented clade, as has been previously described (157).

Analysis of the JC28 genome with antiSMASH revealed a total of 18 predicted secondary metabolite gene clusters. This included 9 non-ribosomal peptide synthetases (NRPS) clusters, 4 hybrid type 1 polyketide -NRPS (T1PKS-NRPS) clusters, 1 hybrid ladderane-NRPS cluster, 3 bacteriocin clusters, and 1 thiopeptide cluster (**Table S3**). Two of the predicted NRPS clusters in the JC28 genome were found at the end of a contig and five NRPS clusters encompassed an entire contig (**Table S3**), suggesting that the number of clusters may be inflated due to breaks in the assembly at these regions.

The predicted ladderane-NRPS cluster had 100% amino acid similarity to a known bromoalterochromide (BAC) producing cluster from *Pseudoalteromoans piscicida* JCM20779 (116). This operon contains 14 genes, including three NRPS genes (*altKLM*) and a halogenase (*altN*) which attaches the bromine to the alterochromide (**Fig. 3B, Table 2**). Bromoalterochromides are yellow lipopeptides known to have cytotoxic activity on sea urchin eggs (112) and the ciliate, *Tetrahymena pyriformis* (114), and antibacterial activity against *Bacillus subtilis* (116). Within the pigmented tree group, all strains in the branch that contained JC28 also contained the BAC biosynthetic gene cluster (**Table S3**). Of the other bioactive secondary metabolites produced by pseudoalteromonads with known biosynthetic pathways, *P.* sp. JC28 also has the gene for production of the antibacterial peptide, Pfap, produced by *P.*

flavipulchra JG1 (139) (99% nucleotide similarity). We did not detect genes for bromopyrroles/bromophenols, tambjamines, thiomarinols, or violacein.

Isolation and structure elucidation of bromoalterochromides (BACs)

The organic extracts obtained from both the supernatant and pellet of the JC28 culture were analyzed by liquid chromatography-mass spectrometry (LC-MS). The EtOAc extract from the supernatant, as well as the 90% aqueous MeOH extract from the pellet, showed several peaks that strongly absorbed at 395 nm, and with MS spectra that were characteristic of brominated compounds (**Fig. S1, Table S5**). These two extracts were then purified separately by reversed-phase high performance liquid chromatography (RP-HPLC), resulting in a combined total of 4.1 mg of **1a**/**1b** and 1.1 mg of **2a**/**2b**. Extensive analysis of the isolated compounds by NMR spectroscopy revealed that **1a**/**1b** is a mixture of constitutional isomers of known BACs (**Fig. 3A**), BAC-A/A'. The spectroscopic data were superimposable with those in the literature (112). Compounds **2a**/**2b** were elucidated to be another pair of isomers of two new BAC derivatives, we have named BAC-D/D'. The new BAC's have a mass 14 Da greater than that of **1a**/**1b**, which suggested the presence of an additional CH_2 in their structure. Although two studies have previously detected these new derivatives from *P. piscicida* JCM20779 and *P. maricaloris* KMM636 (111, 159), their actual structures have not been reported. The HMBC and NOESY NMR data of **2a** were identical with those of **1a**, indicating that the connectivities between the five amino acid residues and the acyl chain are the same (**Fig. 4**). The ^{1}H NMR spectra of the two pairs of derivatives were almost completely identical, except that the proton signals at 3.96 and 1.87 ppm (**Table 3**), assigned to H-2 and H-3 of the Val moiety in **1a**/**1b**, could not be observed in the ^{1}H NMR spectrum of **2a**/**2b**. This clearly suggested that the Val residue should

be replaced with a different amino acid, possibly Ile or Leu, in **2a**/**2b**. COSY correlations between a CH proton at 1.68 ppm and doublet CH_3 protons at 0.90 ppm, and CH_2 protons at 1.13 and 1.42 ppm confirmed that **2a**/**2b** possess an Ile moiety in place of Val; hence, the structure of **2a**/**2b** was determined (**Fig. 3A**).

Besides this, our group noted necessary chemical shift reassignments on the aromatic ring carbons (C-1', C-2', and C-6') of BAC-A/A' because they were incongruent with the previous assignments reported by Speitling et al. (112)(**Table 3**). Moreover, because of the higher spectroscopic resolution used in this study, which was further enhanced by a cryoprobe, the chemical shifts for all the olefins were definitively assigned. High-resolution MS analysis of **1a**/**1b** and **2a**/**2b** showed *m/z*'s of 844.2899 Da and 858.3034 Da, respectively, which are in full agreement with the calculated *m/z*'s for the respective BAC's $[M+H]^+$, 844.2875 Da and 858.3032 Da.

We set out to confirm the absolute configuration of the isolated compounds because of the inconsistencies found in the literature. Speitling et al. (112) tentatively assigned all five amino acids in the peptide portion of BAC-A/A' to be of L-configuration. Furthermore, their group, as well as Kalinovskaya et al. (113), assigned Thr to be in the *threo*-form. These stereochemical assignments are almost the exact opposite of the bioinformatic data obtained from the *P. piscicida* JCM20779 and JC28 BAC gene clusters, wherein all nonribosomal peptide synthetase (NRPS) modules bear an epimerase domain, except for that of Asn-1, and the module for Thr was predicted to favor its *erythro*-diastereomer (116, 159). To verify the absolute stereochemistry, the C3 Marfey method was performed on the acid hydrolysates of **1a**/**1b** and **2a**/**2b**, as well as on all isomers of the standard amino acids (118). In addition, since Asn was expected to be converted, in part or completely, to Asp under the hydrolytic conditions applied to

the natural products (160), standards of Asp were derivatized and analyzed in the same manner. Analyzing the L-FDLA-derived samples by LC-MS condition A, **1a**/**1b** were found to bear D-Val, L- and D-Leu (1:6), D-Ile or D-aIle, D-aThr, and L- and D-Asp (1:2) (**Table 4**). No peak corresponded to Asn, indicating complete hydrolysis to Asp. Since both isomers of Asp were detected from the hydrolysate, Asn-1 and Asn-2 were assigned to be in the L-form and D-form, respectively, in accordance with the JC28 bioinformatic data. However, under the employed analytical conditions, D-Ile and D-aIle could not be fully resolved. Consequently, L-FDAA-derived samples were prepared and analyzed using LC-MS condition B, where D-aIle, rather than D-Ile, was observed in the **1a**/**1b** hydrolysate. Similar results were acquired for **2a**/**2b**, with the exception that a Val peak was not detected. Therefore, the complete structures were designated as D-aThr-D-Val-L-Asn1- D-Asn2-D-aIle/D-Leu for **1a**/**1b**, and D-aThr-D-aIle-L-Asn1-D-Asn2- D-aIle/D-Leu for **2a**/**2b**.

Antibacterial and NO inhibition assays

To investigate the potential role of these PKS-NRPS metabolites in JC28's antimicrobial activity, BAC-A/A' and BAC-D/D' were tested against marine microbes. Due to the scarcity of the available extracts, only one Gram positive strain (*B. algicola* CNJ 803) and one Gram negative strain (*V. fischeri* ES114) were tested. BAC-A/A' (**1a/1b**) had an MIC of 7.4 μM against *B. algicola* and 59.3 μM against *V. fischeri*. BAC-D/D' (**2a/2b**) had an MIC of 7.3 μM against *B. algicola* and >58.3 μM against *V. fischeri*. Interestingly, both BAC samples were 8-fold more active against the *Bacillus* than the *Vibrio*. Furthermore, **1a**/**1b** were subjected to NO inhibition assay to demonstrate a possible mechanism for host colonization and how the symbiont escapes

the host immune response. As a result, co-incubation of LPS-treated RAW264.7 macrophage cells with **1a**/**1b** resulted in inhibition of NO production (IC_{50} 16.9 μM).

Antifungal and cytotoxicity assays

Since the JC28 isolate strongly reduced fungal hyphal growth of *Fusarium keratoplasticum* FSSC-2g by 91% in a well diffusion assay (Appendix I), **1a**/**1b** was tested in a similar assay to confirm whether the BAC's may be, at least, partly responsible for the observed activity. When 50 μg of **1a**/**1b** was added onto a 6 mm paper disc inoculated with the fungus, a 28% reduction in the surface area of hyphal growth was observed, in comparison with the negative control, after 72 h of incubation (**Figure S2**). In addition, **1a**/**1b** was subjected to a cytotoxicity assay versus MCF-7 cells, wherein the compounds demonstrated strong activity.

Discussion

The *Pseudoalteromonas* genus is comprised of Gram negative, heterotrophic, and aerobic marine bacteria that belong to the sub-class of *Gammaproteobacteria* (157, 161, 162), and which comprises less than 0.1% of the *E. scolopes* ANG/JC bacterial communities (72). The pseudoalteromonads demonstrate a broad range of antimicrobial, anti-settlement, and cytotoxic activity on many organisms. This activity is thought to be attributed mainly to the pigmented members of this group (157), to which *Pseudoalteromonas* sp. JC28 belongs and has demonstrated antibacterial activity here and previous antifungal activity (Appendix I). Within the branch containing JC28, many *P. piscicida* strains have also shown strong inhibition against vibrios and other bacterial pathogens (156)(163) and antifungal activity against several fungi, including *Arthrinium* c.f. *saccharicola* (164) and *Fusarium oxysporum* (165). *P. piscicida* strains

have demonstrated eukaryotic toxicity, killing *Artemia nauplii* and *C. elegans* nematodes (136), and lysing *Gymnodinium catenatum* algal cells (166). Other *P. piscicida* strains inhibited settlement of red algae (*Polysiphonia* sp.) and green algae (*Ulva lactuca*) spores (156)(150).

The bacterial community in the jelly coat of *E. scolopes* eggs is necessary for preventing fouling by bacteria and fungi, which would otherwise result in embryo death (Appendix I). The exact cause of embryo death in fouled eggs is unclear, but it may result from a decrease in oxygen. Oxygen levels in marine gastropod egg clutches can drop from 100% of air saturation to 50% at 5 mm inside a clutch during early developmental stages, and drop to nearly zero in late development stages (24). Any added biomass (microbes, algae, or settled invertebrate larva) on the exterior of egg clutches could further deplete available oxygen to the embryos. Outside of the branch containing JC28 (**Fig. 2**), the biofilms of several *Pseudoalteromonas* species are actually necessary for settlement or induction of morphogenesis of marine invertebrate larvae, such as the polychaete *Hydroides elegans* (167), the coral *Acropora millepora* (168), or the sea urchin *Heliocidaris erythrogramma* (169). Although JC28 does not have the settlement genes associated with *P. luteoviolacea* HI1 (170), future research will need to examine what impact JC28 has on larval or algal settlement. Overall, the effect of clearing the JC bacteria from *E. scolopes* eggs and determining the effect of larval/algal settlement on embryo development remains to be tested.

The closest relative of the JC28 strain, *P. flavipulchra* JG1, was isolated from the rearing water of healthy turbot in Qingdao, China, and was investigated as a potential probiotic in fish rearing (171). It strongly inhibited *Aeromonas* and *Vibrio* species *in vitro*, and did not have any eukaryotic toxicity against zebra fish, bivalves, or mantis shrimp (171). Since JC28 was isolated from healthy *E. scolopes* eggs, and was presumably deposited into these eggs from the ANG

community of adult female squid, it likely also does not have toxic effects on its host. The antibacterial activity of JG1 was strongest against *V. anguillarum in vitro*, similar to what was seen here with JC28 (139). Several antibacterial compounds have been characterized from JG1, including the Pfap protein and five small molecules: *p*-hydroxybenzoic acid, *trans*-cinnamic acid, 6-bromoindolyl-3-acetic acid, *N*-hydroxybenzoisoxazolone, and 2'-deoxydenosine, with the brominated compound (6-bromoindolyl-3-acetic acid) having the most broad antimicrobial activity (139). The Pfap protein is a putative L-amino acid oxidase, which may play a role in synthesis of 6-bromoindolyl-3-acetic acid (172). Future research will need to determine if Pfap, also contributes to JC28's inhibitory activity. Although JG1's bromoalterochromide activity is not described, we tested the activity of JC28's BACs here.

The BACs produced by JC28 demonstrated increased antibacterial activity against a Gram positive bacterium over a Gram negative bacterium, and had activity against the fungal pathogen *F. keratoplasticum.* The BACs did demonstrate a large gap in antifungal activity (28% hyphal inhibition) compared to the whole isolate's *in vitro* activity (91% hyphal inhibition, Appendix I). A similar observation has been described in other BAC-producing *Pseudoalteromonas* strains, which were initially active against *Vibrio anguillarium* in an isolate well-diffusion assay, but whose EtOAc extracts turned out to be inactive (115). This could be caused by upregulation in BAC production when exposed to a pathogen, or that water-soluble metabolites in the cell-free supernatant are necessary to exert maximum hyphal growth inhibition. Although BAC-A isolated from *Pseudoalteromonas* sp. J010 was previously shown to have antiprotozoal activity against *Tetrahymena pyriformis*, it did not have antibacterial or antifungal activity in a disc-diffusion assay (114). This is the first demonstration that purified BACs themselves, devoid of the producer strain or other bacterial metabolites, possess

antibacterial and antifungal activity, signifying the ecological and clinical importance of these PKS-NRPS molecules. The cytotoxic activity of BAC-A/A' versus MCF-7 human breast cancer cells, but not against RAW264.7 macrophage cells in the NO inhibition assay (data not shown), suggests selectivity toward tumor cell lines and makes this noteworthy for drug discovery. Heterologous expression of the BAC gene cluster in *E. coli* (159) and optimization of *Pseudoalteromonas* culturing conditions (111) are promising methods to obtain higher yields of BACs for further biological investigations.

Besides potentially aiding in *E. scolopes* egg defense, the BACs may play a role in *Pseudoalteromonas* sp. JC28's ability to colonize the ANG. In the well-studied light organ symbiosis of *E. scolopes*, the bioluminescent *V. fischeri* exclusively colonizes the light organ just a few hours after the eggs hatch (173–175). Every morning, 95% of the *V. fischeri* population is expelled from the light organ, as part of a daily venting routine (176, 177). Proteomic analysis of this exudate, along with surrounding host epithelial tissue, revealed that the host expresses several proteins involved in oxidative stress, resulting in production of reactive oxygen and nitrogen species (ROS and NO). In response to the host, symbiont proteins responsible for ROS and NO detoxification, such as superoxide dismutase and nitric oxide dioxygenase, are expressed and detected in the light organ exudate (173). The ability of BACs to inhibit NO production in the *in vitro* assay using LPS-induced RAW264.7 macrophage cells may be an indication of how JC28 may overcome the host's immune system to colonize the ANG, and eventually the JC. Further investigations are necessary, however, to determine whether the ANG colonization process follows the same principles of the light organ model.

Conclusions

Pseudoalteromonas sp. JC28 represents the first sequenced *Gammaproteobacteria* isolate of the *E.scolopes* ANG symbiosis and demonstrates broad antimicrobial activity against Gram positive and Gram negative bacteria, as well as fungi. Although pseudoalteromonads make up less than 0.1% of the ANG/JC bacterial community (72), its strong bioactivity may contribute to the community's overall defensive function. *Pseudoalteromonas* sp. JC28 has been shown to produce bromoalterochromides with antimicrobial and select cytotoxic effects, making this a potential contributor to JC28's bioactivity. Future research will need to knockout the bromoalterochromide biosynthetic genes, and the genes for other predicted secondary metabolites, to determine the role of these compounds in JC28's inhibitory activity, and potential function in the ANG symbiosis.

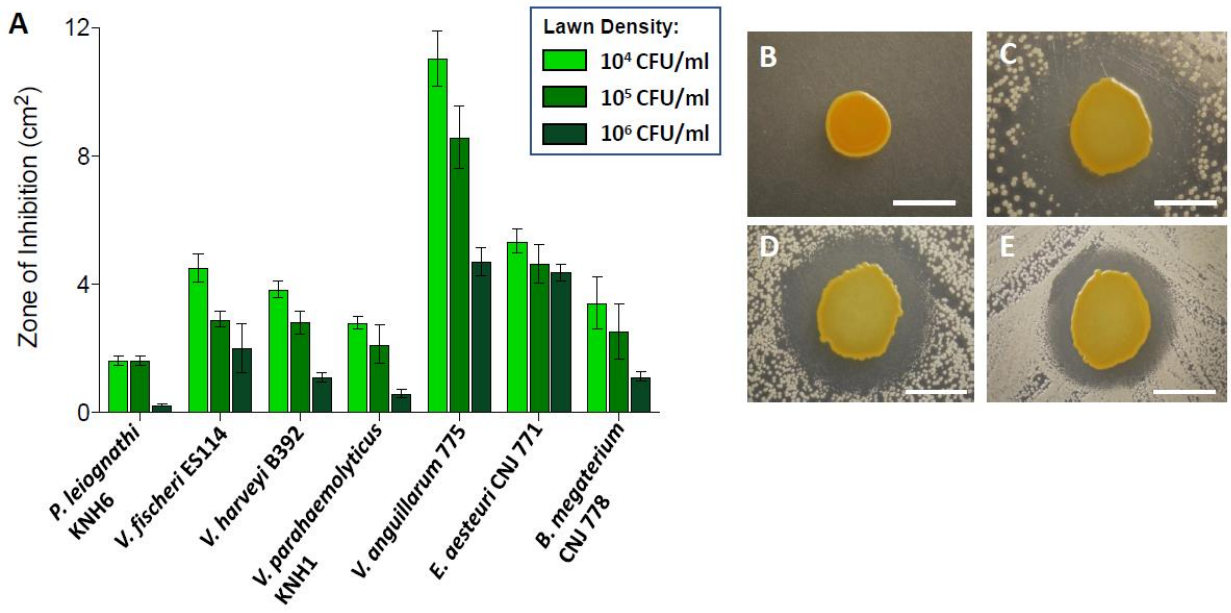


Figure 1. *Pseudoalteromonas* sp. JC28 inhibits Gram positive and Gram negative marine bacteria. (A) Areas of inhibition (cm^2) were measured on lawns of marine bacteria with increasing lawn densities. Significantly greater inhibition was noted between 10^4 – 10^5 and 10^5 – 10^6 CFU/ml densities of all species with the exception of the 10^4 – 10^5 CFU/ml densities of *P. leiognathi* and the 10^5 – 10^6 CFU/ml densities of *E. aesteuri* (post hoc Tukey tests from two-way ANOVA). Data presented as mean of nine trials +/- standard deviation. Representative images of **(B)** JC28 monoculture control, and JC28 cocultured with *P. leiognathi* at **(C)** 10^4 CFU/ml, **(D)** 10^5 CFU/ml, and **(E)** 10^6 CFU/ml lawn densities. Scale bars = 5 mm.

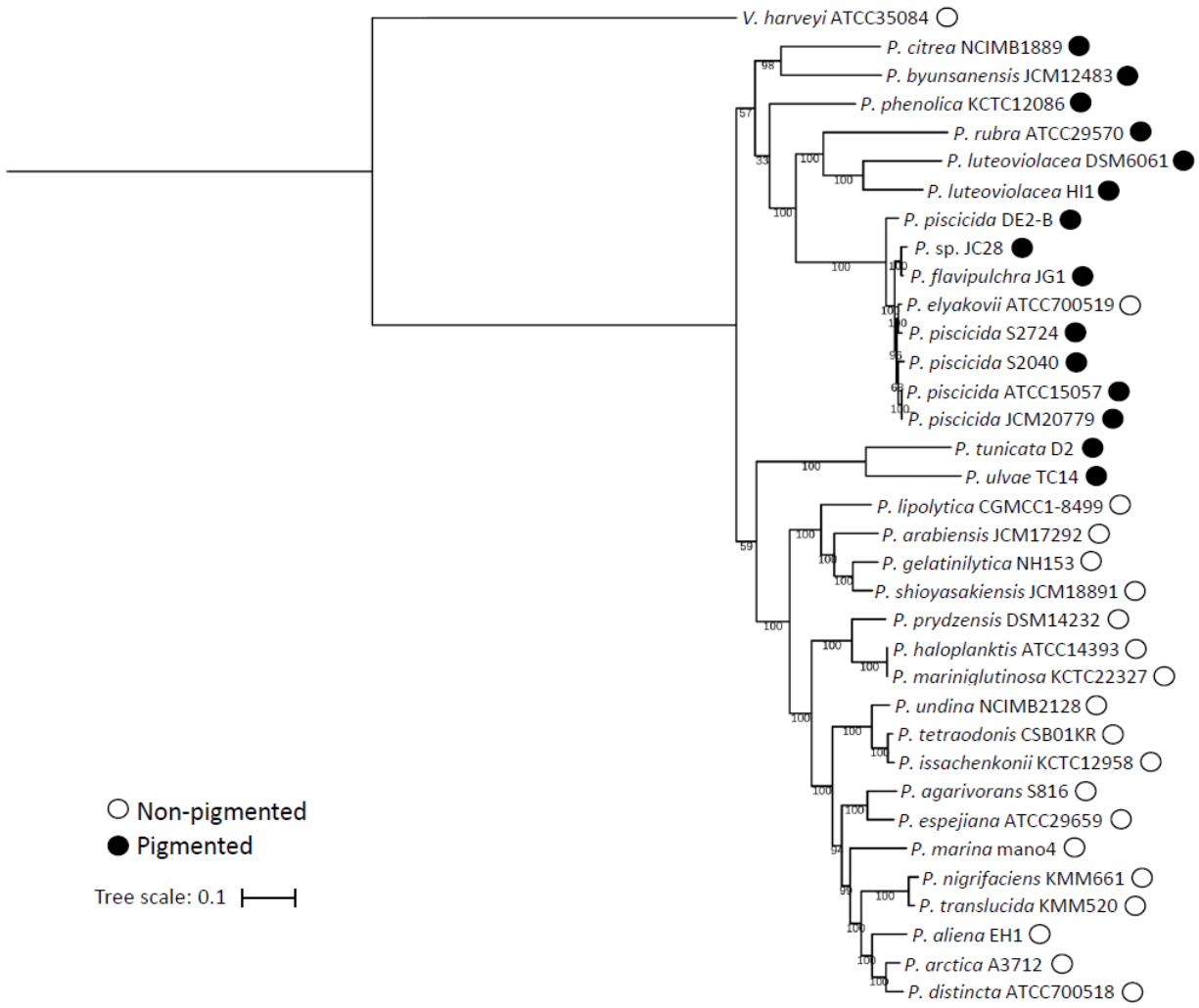


Figure 2. MLSA analysis of *Pseudoalteromonas* genomes and isolate from *E. scolopes* egg jelly coat. Phylogenetic analysis was performed with 6 single-copy, housekeeping genes. Isolate JC28 falls within a clade containing the pigmented *P. flavipulchra* JG1. Black circles indicate pigmented strains, white circles indicate non-pigmented strains.

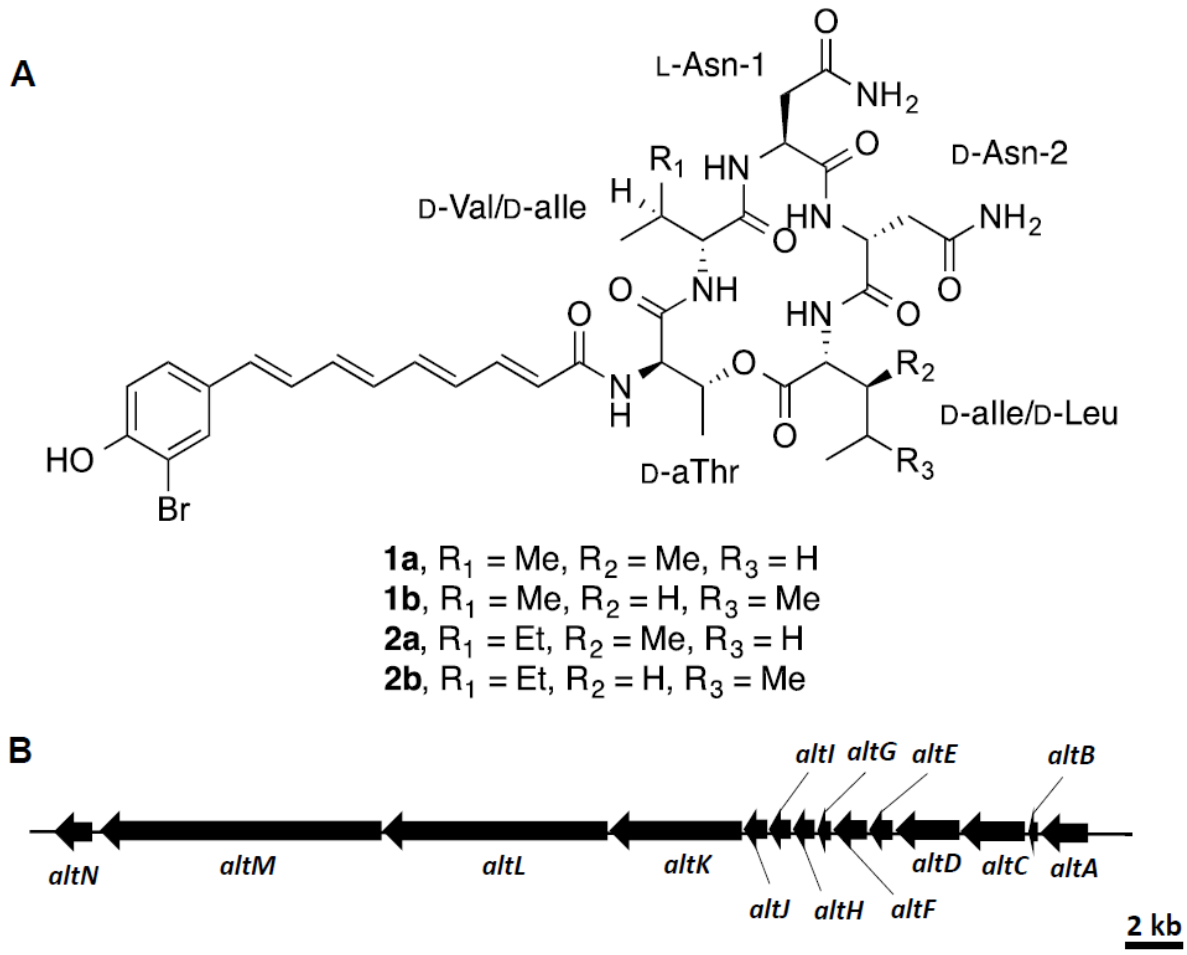


Figure 3. Structures and biosynthetic operon of bromoalterochromides. (A) Structures of bromoalterochromides A/A' (**1a**/**1b**) and D/D' (**2a**/**2b**). **(B)** Bromoalterochromide biosynthetic operon.

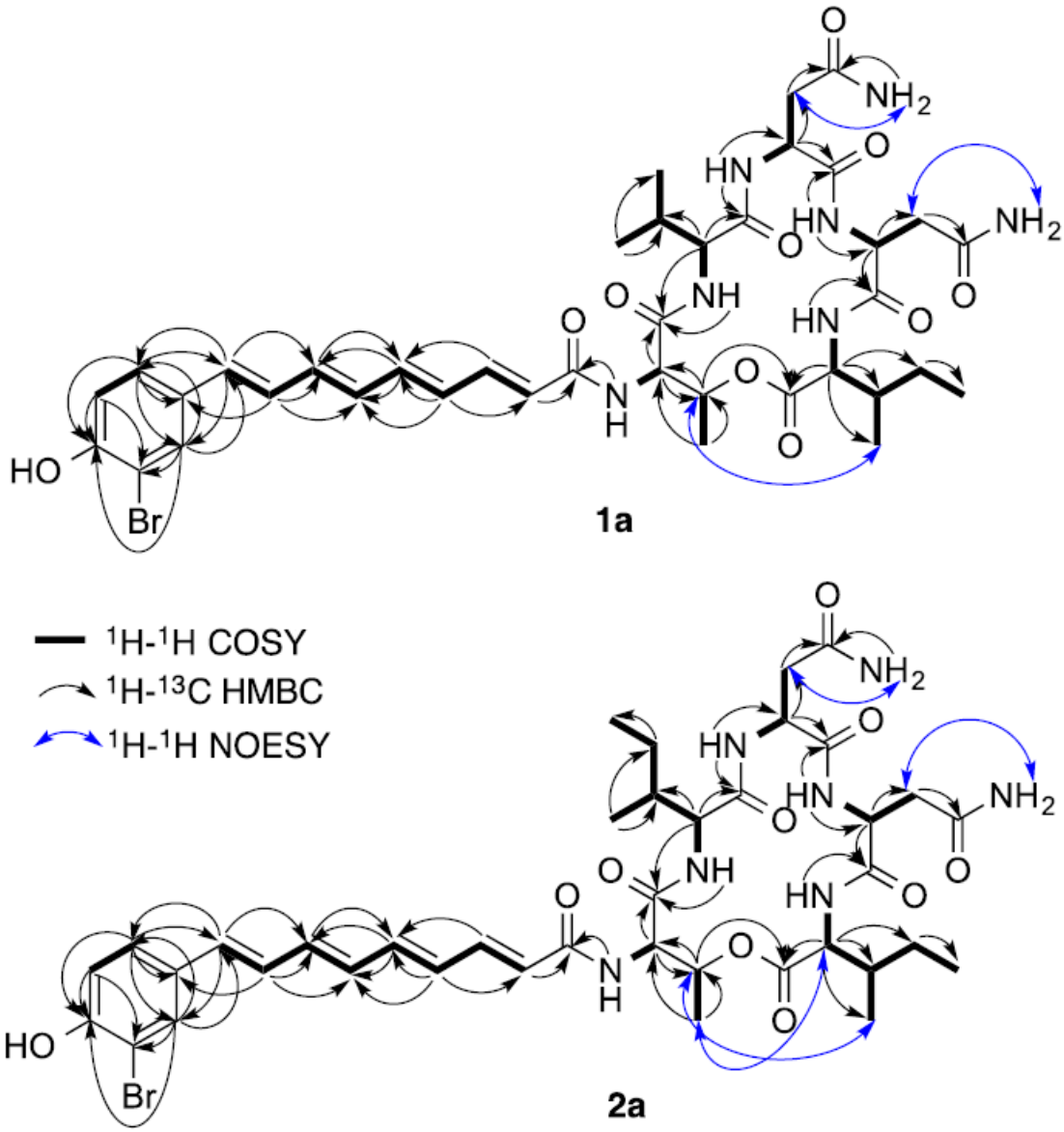


Figure 4. COSY, HMBC, and NOESY correlations in **1a** and **2a**.

Table 1. Summary of ANG/JC isolates that showed inhibition against seven marine bacteria *in vitro*

	ANG/JC Isolate	1[a]	2	3	4	5	6	7
Alphaproteobacteria	*Leisingera* sp. ANG1	-	-	-	-	-	-	-
	Leisingera sp. ANG-M6	-	-	-	-	-	-	-
	Leisingera sp. ANG-DT	-	-	-	-	-	-	-
	Leisingera sp. ANG-S	-	-	-	-	-	-	-
	Leisingera sp. ANG-S3	-	-	-	-	-	-	-
	Leisingera sp. ANG52	-	-	-	-	-	-	-
	Leisingera sp. ANG59	+	+	-	-	-	-	-
	Leisingera sp. JC11	-	-	-	-	-	-	-
	Rhodobacteraceae sp. ANG7	-	-	-	-	-	-	-
	Rhodobacteraceae sp. ANG13	-	-	-	-	-	-	-
	Labrenzia sp. ANG18	-	-	-	-	-	-	-
	Ruegeria sp. ANG6	-	-	-	-	-	-	-
	Ruegeria sp. ANG10	-	-	-	-	-	-	-
	Ruegeria sp. ANG-S4	-	-	-	-	-	-	-
Flavo.	*Muricauda* sp. ANG21	-	-	-	-	-	-	-
Gamma.	*Shewanella* sp. ANG44	-	-	-	-	-	-	-
	Altermonas sp. JC21	-	-	-	-	-	-	-
	Pseudoalteromonas sp. JC28	+	+	+	+	+	+	+
	Vibrio sp. JC34	+	-	-	-	-	-	-

Abbreviations: *Flavo. = Flavobacterriia*; *Gamma. = Gammaproteobacteria*; + = zone of inhibition observed, - = no zone of inhibition observed (results of 3 trials).

[a]Numbers refer to the following strains: 1, *Vibrio fischeri* ES114; 2, *Vibrio anguillarum* 775; 3, *Vibrio parahaemolyticus* KNH1; 4, *Vibrio harveyi* B392; 5, *Photobacterium leiognathid* KNH6; 6, *Bacillus megaterium* CNJ778; 7, *Exiguobacterium aestuarii* CNJ771

Table 2. Comparison of the bromoalterochromide biosynthetic gene cluster in *Pseduoalteromonas* sp. JC28 and *Pseudoalteromoans piscicida* JCM20779

Gene	Gene function	*P.* sp. JC28 locus ID	*P. piscicida* locus ID	% Amino acid similarity to *P. piscicida* operon	E-value	Percent query cover
altN	Flavin-dependent halogenase	DS891_00345	PpisJ2_10100002431	97.82	0	100
altM	NRPS	DS891_00350	PpisJ2_10100002426	96.89	0	100
altL	NRPS	DS891_00355	PpisJ2_10100002421	97.28	0	100
altK	NRPS	DS891_00360	PpisJ2_10100002416	98.42	0	100
altJ	Thioesterase	DS891_00365	PpisJ2_10100002411	98.01	0	100
altI	Hypothetical protein	DS891_00370	PpisJ2_10100002406	95.74	9e-171	100
altH	SDR family NAD(P)-dependent oxidoreductase	DS891_00375	PpisJ2_10100002401	99.18	0	100
altG	3-hydroxydecanoyl-ACP dehydratase	DS891_00380	PpisJ2_10100002396	100	6e-99	100
altF	ABC transporter permease	DS891_00385	PpisJ2_10100002391	100	0	100
altE	ABC transporter ATP-binding protein	DS891_00390	PpisJ2_10100002386	100	0	100
altD	3-oxoacyl-ACP synthase	DS891_00395	PpisJ2_10100002381	97.66	0	100
altC	acyl-coA synthetase/AMP-fatty acid ligase	DS891_00400	PpisJ2_10100002376	97.99	0	100
altB	Acyl carrier protein	DS891_00405	PpisJ2_10100002371	100	3e-68	100
altA	Aromatic amino acid lyase	DS891_00410	PpisJ2_10100002366	99.63	0	100

*Abbreviations: NRPS = non-ribosomal peptide synthetase

Table 3. ^{1}H and ^{13}C chemical shifts of **1a**/**1b** and **2a**/**2b** in DMSO-d_6 (600 MHz).

	1a		**1b**		**2a**		**2b**	
	δ_H, multiplicity ($^3J_{H-H}$)	δ_C	δ_H, multiplicity ($^3J_{H-H}$)	δ_C	δ_H, multiplicity ($^3J_{H-H}$)	δ_C	δ_H, multiplicity ($^3J_{H-H}$)	δ_C
D-Val/aIle-1								
NH	8.18, d (6.0)	--	8.18, d (6.0)	--	8.11, m	--	8.11, m	--
1	--	171.9	--	171.9	--	172.2	--	172.2
2	3.96, m	59.0	3.96, m	59.0	4.18, m	57.1	4.18, m	57.1
3	1.87, m	29.5	1.87, m	29.5	1.68, m	35.8	1.68, m	35.8
3-Me	0.88, d (6.0)	18.6	0.88, d (6.0)	18.6	0.90, d (6.0)	14.8	0.90, d (6.0)	14.8
4	0.95, d (6.0)	18.9	0.95, d (6.0)	18.9	1.13, m 1.43, m	25.6	1.13, m 1.43, m	25.6
5	--	--	--	--	0.84, m	11.3	0.84, m	11.3
L-Asn-1								
NH	8.65, d (6.0)	--	8.70, d (6.0)	--	8.64, d (6.0)	--	8.67, d (6.0)	--
1	--	171.4	--	170.2	--	171.4	--	170.3
2	4.32, m	51.1	4.30, m	51.1	4.30, m	51.7	4.28, m	51.7
3	2.41, m	35.8	2.41, m	35.8	2.41, m	36.1	2.41, m	36.1
4	--	170.7	--	170.7	--	170.7	--	170.7
NH_2	6.90, br 7.28, br		6.90, br 7.28, br		6.91, br 7.30, br		6.91, br 7.30, br	
D-Asn-2								
NH	8.11, br	--	7.70, d (6.0)	--	8.12, m	--	7.73, d (6.0)	--
1	--	170.0	--	170.2	--	170.0	--	170.3
2	4.39, m	50.6	4.43, m	50.6	4.39, m	50.7	4.43, m	50.7
3	2.74, dd (18.0,6.0)	35.1	2.55, m	35.1	2.75, dd (18.0,6.0)	35.2	2.55, m	36.0
4	--	172.2	--	170.2	--	172.2	--	172.2
NH_2	6.82, br 7.25, br	--	6.82, br 7.25, br	--	6.82, br 7.30, br	--	6.82, br 7.30, br	--
D-aIle/Leu								
NH	7.15, m	--	8.36, d (6.0)	--	7.22, m	--	8.35, d (6.0)	--
1	--	169.3	--	171.1	--	169.3	--	171.1
2	4.36, m	55.8	4.43, m	49.7	4.35, m	56.4	4.45, m	49.7
3	1.96, m	36.4	1.65, m 1.88, m	38.7	1.99, m	36.6	1.64, m 1.88, m	38.4

3-Me/4-Me	0.80, d (6.0)	14.3	0.84, m	21.4	0.80, d (6.0)	14.6	0.84, m	21.9
4	1.05, m 1.33, m	25.7	1.63, m	23.7	1.05, m 1.33, m	25.7	1.63, m	23.7
5	0.89, t (6.0)	11.1	0.91, d (6.0)	23.7	0.89, t (6.0)	11.3	0.91, d (6.0)	23.7
D-aThr								
NH	8.25, d (6.0)	--	8.25, d (6.0)	--	8.28, m	--	8.28, m	--
1	--	168.8	--	169.2	--	168.8	--	168.8
2	4.74, m	54.8	4.82, m	54.8	4.77, m	55.0	4.77, m	55.0
3	4.75, m	72.2	4.75, m	72.2	4.76, m	72.3	4.76, m	72.3
3-Me	1.36, d (6.0)	16.6	1.36, d (6.0)	16.6	1.34, d (6.0)	16.5	1.34, d (6.0)	16.5
Acyl								
1	--	164.7	--	164.7	--	164.8	--	164.8
2	6.19, d (12.0)	123.7	6.19, d (12.0)	123.7	6.20, d (12.0)	123.8	6.20, d (12.0)	123.8
3	7.14, m	139.6	7.14, m	139.6	7.13, m	139.6	7.13, m	139.6
4	6.42, m	129.9	6.42, m	129.9	6.40, m	129.9	6.40, m	129.9
5	6.72, dd (18.0, 12.0)	139.1	6.72, dd (18.0, 12.0)	139.1	6.72, dd (18.0, 12.0)	139.3	6.72, dd (18.0, 12.0)	139.3
6	6.47, m	131.7	6.47, m	131.7	6.45, m	131.5	6.45, m	131.5
7	6.56, dd (18.0, 12.0)	136.2	6.56, dd (18.0, 12.0)	136.2	6.54, m	136.4	6.54, m	136.4
8	6.86, m	126.9	6.86, m	126.9	6.89, m	126.9	6.89, m	126.9
9	6.58, d (18.0)	132.6	6.58, d (18.0)	132.6	6.58, m	132.4	6.58, m	132.4
1'	--	129.8	--	129.8	--	129.7	--	129.7
2'	7.65, s	130.6	7.65, s	130.6	7.64, s	130.6	7.64, s	130.6
3'	--	109.8	--	109.8	--	109.8	--	109.8
4'	--	153.9	--	153.9	--	154.2	--	154.2
4'-OH	nd	--	nd	--	nd	--	nd	--
5'	6.90, m	116.3	6.90, m	116.3	6.92, m	116.3	6.90, m	116.3
6'	7.32, m	126.7	7.32, m	126.7	7.32, m	126.7	7.32, m	126.7

nd, not detected

Table 4. Retention times (t_R) of standards and hydrolysate derivatives using L-FDLA and L-FDAA.

	LC-MS Condition	t_R (min)		
		Standards	**1a/1b**	**2a/2b**
L-Val	A	30.01	nd	nd
D-Val		37.09	37.12	nd
L-Leu	A	31.55	31.57	nd
D-Leu		40.10	40.13	40.58
L-Ile	A	31.37	nd	nd
D-Ile		39.27	39.27	39.66
L-aIle		30.81	nd	nd
D-aIle		39.21	39.27	39.66
L-Thr	A	21.70	nd	nd
D-Thr		28.52	nd	nd
L-aThr		23.42	nd	nd
D-aThr		25.95	25.96	26.26
L-Asp	A	23.97	23.97	24.22
D-Asp		27.16	27.13	27.54
D-Ile	B	73.64	nd	nd
D-aIle		74.33	74.67	74.80

nd, not detected

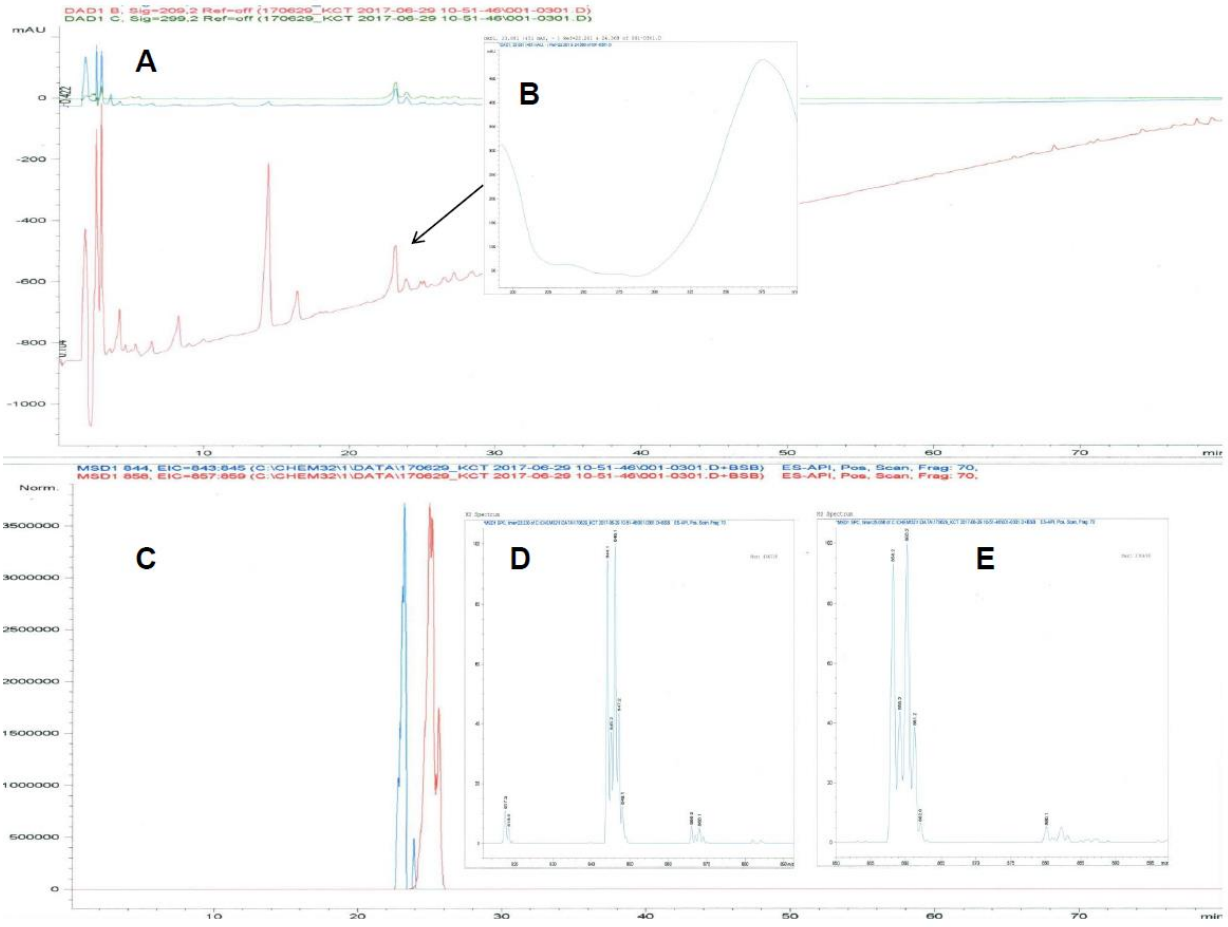


Figure S1. Bromoalterochromide chromatograms. (A) LC chromatogram of BAC-containing EtOAc extract from JC28 culture supernatant, ran using a long gradient, monitored at 209 nm (red trace), 254 nm (blue trace), and 299 nm (green trace). (B) UV chromatogram of peak eluting at 23.08 min. (C) Extracted ion chromatogram at m/z 844 (blue trace) and 858 (red trace). (D) MS spectrum of peak at 23.23 min. (E) MS spectrum of peak at 25.06 min.

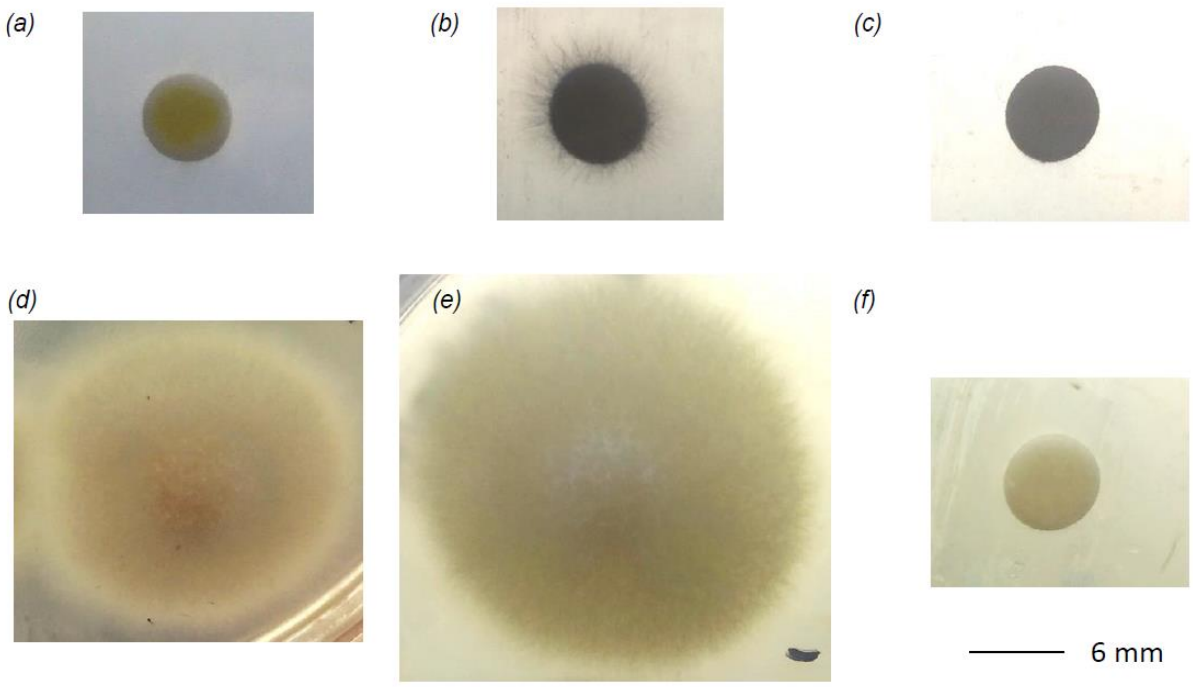


Figure S2. Bromoalterochromides inhibited hyphal growth of F. keratoplasticum-2g in the modified disc diffusion assay. (a)-(c) 24 h, and (d)-(f) 72 h after incubation. (a) and (d) 1a/1b (50-μg), (b) and (e) negative control, DMSO (10 μL), (c) and (f) positive control, cycloheximide (68 μg). A 28% reduction in the area of hyphal growth was observed in (d), relative to (e).

Table S1. Bacterial strains used in this study.

Strain	Characteristics	Source
Leisingera sp. ANG1	Isolate from *Euprymna scolopes* ANG	Collins & Nyholm (2011)
Ruegeria sp. ANG6	Isolate from *E. scolopes* ANG	This study
Leisingera sp. ANG7	Isolate from *E. scolopes* ANG	This study
Ruegeria sp. ANG10	Isolate from *E. scolopes* ANG	This study
Leisingera sp. ANG13	Isolate from *E. scolopes* ANG	This study
Labrenzia sp. ANG18	Isolate from *E. scolopes* ANG	This study
Muricauda sp. ANG21	Isolate from *E. scolopes* ANG	Gromek et al. (2016)
Shewanella sp. ANG44	Isolate from *E. scolopes* ANG	This study
Leisingera sp. ANG52	Isolate from *E. scolopes* ANG	This study
Leisingera sp. ANG59	Isolate from *E. scolopes* ANG	This study
Leisingera sp. ANG-DT	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-M6	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-S	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. ANG-S3	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Ruegeria sp. ANG-S4	Isolate from *E. scolopes* ANG	Collins et al. (2015)
Leisingera sp. JC11	Isolate from *E. scolopes* egg jelly coats	This study
Alteromonas sp. JC21	Isolate from *E. scolopes* egg jelly coats	This study
Pseudoalteromonas sp. JC28	Isolate from *E. scolopes* egg jelly coats	This study
Vibrio sp. JC34	Isolate from *E. scolopes* egg jelly coats	This study
Photobacterium leiognathi KNH6	Seawater isolate	Stabb & Ruby (2002)
Vibrio anguillarum 775	Isolate from *Oncorhynchus kisutch*	Crosa et al. (1977) (Courtesy of Dr. Joerg Graf, University of Connecticut)
Vibrio fischeri ES114	Isolate from *E. scolopes* light organ	Boettcher & Ruby (1990)
Vibrio harveyi B392	Seawater isolate	Reichelt & Baumann (1973)
Vibrio anguillarum 775	Isolate from *Oncorhynchus kisutch*	Crosa et al. (1977) (Courtesy of Dr. Joerg Graf, University of Connecticut)
Bacillus algicola CNJ 803	Marine sediment sample, 0 meter depth, Republic of Palau	Gontang et al. (2007) (Courtesy of Dr. René Augustin and the McFall-Ngai/Ruby Labs, University of Hawaii)
Bacillus megaterium CNJ 778	Marine sediment sample, 23 meter depth, Republic of Palau	Gontang et al. (2007) (Courtesy of Dr. René Augustin and the McFall-

		Ngai/Ruby Labs, University of Hawaii)
Exiguobacterium aestuarii CNJ 771	Marine sediment sample, 50 meter depth, Republic of Palau	Gontang et al. (2007) (Courtesy of Dr. René Augustin and the McFall-Ngai/Ruby Labs, University of Hawaii)

Table S2. Zone of inhibition areas of *Pseudoalteromonas* sp. JC28 assayed against the target bacteria

Target strain	**10^4 lawn density average area (cm^2)**	**10^5 lawn density average area (cm^2)**	**10^6 lawn density average area (cm^2)**
P. leiognathi KNH6	1.63 ± 0.05	1.63 ± 0.05	0.24 ± 0.01
V. harveyi B392	3.85 ± 0.09	2.82 ± 0.12	1.10 ± 0.05
V. fischeri ES114	4.50 ± 0.15	2.91 ± 0.08	2.02 ± 0.25
V. parahaemolyticus KNH1	2.80 ± 0.07	2.14 ± 0.20	0.59 ± 0.04
V. anguillarum 775	11.1 ± 0.29	8.59 ± 0.33	4.72 ± 0.14
E. aesteuri CNJ 771	5.34 ± 0.13	4.65 ± 0.20	4.38 ± 0.08
B. megaterium CNJ 778	3.42 ± 0.27	2.55 ± 0.29	1.12 ± 0.05

Areas are an average of 9 trials ± the standard error of the mean.

Table S3. Bromoalterochromide gene clusters in tree clade

Strain	Gene	Percent a.a. identity	E-value	Percent query cover
P. elyakovii ATCC700519	*altA*	99.1	0	100.0
P. elyakovii ATCC700519	*altB*	100.0	5.13E-50	100.0
P. elyakovii ATCC700519	*altC*	97.2	0	100.0
P. elyakovii ATCC700519	*altD*	98.8	0	100.0
P. elyakovii ATCC700519	*altE*	100.0	5.11E-174	100.0
P. elyakovii ATCC700519	*altF*	100.0	0	100.0
P. elyakovii ATCC700519	*altG*	100.0	3.06E-86	100.0
P. elyakovii ATCC700519	*altH*	99.6	6.09E-149	91.8
P. elyakovii ATCC700519	*altI*	97.0	5.60E-153	100.0
P. elyakovii ATCC700519	*altJ*	98.4	2.78E-149	100.0
P. elyakovii ATCC700519	*altK*	98.7	0	97.8
P. elyakovii ATCC700519	*altL*	96.7	0	95.0
P. elyakovii ATCC700519	*altM*	98.2	0	79.2
P. elyakovii ATCC700519	*altN*	99.3	0	100.0
P. flavipulchra JG1	*altA*	99.6	0	100.0
P. flavipulchra JG1	*altB*	100.0	5.19E-50	100.0
P. flavipulchra JG1	*altC*	98.2	0	100.0
P. flavipulchra JG1	*altD*	97.8	0	100.0
P. flavipulchra JG1	*altE*	100.0	5.21E-174	100.0
P. flavipulchra JG1	*altF*	100.0	0	100.0
P. flavipulchra JG1	*altG*	100.0	3.10E-86	100.0
P. flavipulchra JG1	*altH*	99.1	1.39E-148	91.8
P. flavipulchra JG1	*altI*	96.2	1.30E-151	100.0
P. flavipulchra JG1	*altJ*	98.4	9.05E-151	100.0
P. flavipulchra JG1	*altK*	98.4	0	97.8
P. flavipulchra JG1	*altL*	97.1	0	88.6
P. flavipulchra JG1	*altM*	97.4	0	69.3
P. flavipulchra JG1	*altN*	99.3	0	100.0
Pseudoalteromonas sp. JC28	*altA*	99.6	0	100.0

Pseudoalteromonas sp. JC28	*altB*	100.0	5.22E-50	100.0
Pseudoalteromonas sp. JC28	*altC*	98.0	0	100.0
Pseudoalteromonas sp. JC28	*altD*	97.7	0	100.0
Pseudoalteromonas sp. JC28	*altE*	100.0	5.24E-174	100.0
Pseudoalteromonas sp. JC28	*altF*	100.0	0	100.0
Pseudoalteromonas sp. JC28	*altG*	100.0	3.12E-86	100.0
Pseudoalteromonas sp. JC28	*altH*	99.1	1.39E-148	91.8
Pseudoalteromonas sp. JC28	*altI*	95.7	9.58E-151	100.0
Pseudoalteromonas sp. JC28	*altJ*	98.0	2.54E-149	100.0
Pseudoalteromonas sp. JC28	*altK*	98.4	0	97.8
Pseudoalteromonas sp. JC28	*altL*	97.3	0	100.0
Pseudoalteromonas sp. JC28	*altM*	96.9	0	100.0
Pseudoalteromonas sp. JC28	*altN*	97.8	0	100.0
P. piscicida ATCC15057	*altA*	100.0	0	100.0
P. piscicida ATCC15057	*altB*	100.0	5.33E-50	100.0
P. piscicida ATCC15057	*altC*	100.0	0	100.0
P. piscicida ATCC15057	*altD*	100.0	0	100.0
P. piscicida ATCC15057	*altE*	100.0	4.99E-174	100.0
P. piscicida ATCC15057	*altF*	100.0	0	100.0
P. piscicida ATCC15057	*altG*	100.0	2.97E-86	100.0
P. piscicida ATCC15057	*altH*	100.0	1.27E-149	91.8
P. piscicida ATCC15057	*altI*	100.0	1.99E-157	100.0
P. piscicida ATCC15057	*altJ*	100.0	3.18E-152	100.0
P. piscicida ATCC15057	*altK*	100.0	0	97.8
P. piscicida ATCC15057	*altL*	99.1	0	98.6
P. piscicida ATCC15057	*altM*	87.5	0	99.0
P. piscicida ATCC15057	*altN*	100.0	0	100.0
P. piscicida DE2B	*altA*	99.1	0	100.0
P. piscicida DE2B	*altB*	100.0	9.52E-50	100.0
P. piscicida DE2B	*altC*	92.8	0	89.0
P. piscicida DE2B	*altD*	96.0	0	100.0

P. piscicida DE2B	*altE*	100.0	5.04E-174	100.0
P. piscicida DE2B	*altF*	100.0	0	100.0
P. piscicida DE2B	*altG*	100.0	2.99E-86	100.0
P. piscicida DE2B	*altH*	98.2	7.07E-147	91.8
P. piscicida DE2B	*altI*	97.0	2.08E-152	100.0
P. piscicida DE2B	*altJ*	95.6	1.19E-146	100.0
P. piscicida DE2B	*altK*	97.4	0	86.8
P. piscicida DE2B	*altL*	96.8	0	100.2
P. piscicida DE2B	*altM*	94.9	0	100.0
P. piscicida DE2B	*altN*	97.3	0	100.0
P. piscicida JCM20779	*altA*	100.0	0	100.0
P. piscicida JCM20779	*altB*	100.0	5.32E-50	100.0
P. piscicida JCM20779	*altC*	100.0	0	100.0
P. piscicida JCM20779	*altD*	100.0	0	100.0
P. piscicida JCM20779	*altE*	100.0	4.94E-174	100.0
P. piscicida JCM20779	*altF*	100.0	0	100.0
P. piscicida JCM20779	*altG*	100.0	2.96E-86	100.0
P. piscicida JCM20779	*altH*	100.0	1.25E-149	91.8
P. piscicida JCM20779	*altI*	100.0	1.97E-157	100.0
P. piscicida JCM20779	*altJ*	100.0	3.14E-152	100.0
P. piscicida JCM20779	*altK*	100.0	0	97.8
P. piscicida JCM20779	*altL*	100.0	0	91.9
P. piscicida JCM20779	*altM*	100.0	0	69.1
P. piscicida JCM20779	*altN*	100.0	0	100.0
P. piscicida S2040	*altA*	99.6	0	100.0
P. piscicida S2040	*altB*	99.0	3.31E-49	100.0
P. piscicida S2040	*altC*	97.6	0	100.0
P. piscicida S2040	*altD*	98.1	0	100.0
P. piscicida S2040	*altE*	100.0	4.68E-174	100.0
P. piscicida S2040	*altF*	100.0	0	100.0
P. piscicida S2040	*altG*	99.2	1.35E-85	100.0

P. piscicida S2040	*altH*	100.0	1.19E-149	91.8
P. piscicida S2040	*altI*	97.9	1.34E-153	100.0
P. piscicida S2040	*altJ*	98.4	1.63E-150	100.0
P. piscicida S2040	*altK*	99.1	0	97.8
P. piscicida S2040	*altL*	98.0	0	80.9
P. piscicida S2040	*altM*	97.5	0	67.6
P. piscicida S2040	*altN*	99.0	0	100.0
P. piscicida S2724	*altA*	99.6	0	100.0
P. piscicida S2724	*altB*	100.0	4.90E-50	100.0
P. piscicida S2724	*altC*	97.8	0	100.0
P. piscicida S2724	*altD*	98.2	0	100.0
P. piscicida S2724	*altE*	99.6	1.93E-173	100.0
P. piscicida S2724	*altF*	99.7	0	100.0
P. piscicida S2724	*altG*	100.0	2.92E-86	100.0
P. piscicida S2724	*altH*	99.1	2.09E-148	91.8
P. piscicida S2724	*altI*	96.2	4.38E-152	100.0
P. piscicida S2724	*altJ*	98.0	2.28E-150	100.0
P. piscicida S2724	*altK*	98.7	0	97.8
P. piscicida S2724	*altL*	96.9	0	100.2
P. piscicida S2724	*altM*	97.7	0	100.0
P. piscicida S2724	*altN*	99.0	0	100.0

Table S4. *Pseudoalteromonas* sp. JC28 antiSMASH results

Cluster type	Most similar known cluster	Scaffold	Total scaffold length (bp)	Cluster bp position	Total cluster length (bp)	% of scaffold containing cluster
Ladderane-NRPS	Alterochromide (100%)	1.1	1,226,393	66,592 – 137,002	70,410	5.7
Bacteriocin	NA	1.1	1,226,393	275,617 – 286,481	10,864	0.89
T1 PKS-NRPS	NA	1.1	1,226,393	633,994 – 683,565	49,571	4
Thiopeptide	NA	1.1	1,226,393	789,245 – 826,848	37,603	3.1
Bacteriocin	NA	3.1	509,001	116,400 – 127,955	11,555	2.3
Bacteriocin	NA	3.1	509,001	209,010 – 220,206	11,196	2.2
NRPS	NA	3.1	509,001	401,470 – 461,115	59,645	11.7
NRPS	NA	6.1	374,113	316,144 – 372,250	56,106	15
T1 PKS-NRPS	Lotilibcin (20%)	7.1	531,598	1 – 86,834	86,834	16.3
T1 PKS-NRPS	NA	8.1	314,808	76,426 – 148,258	71,832	22.8
NRPS	NA	11.1	183,335	159,352 – 183,335	23,983	13.1
T1 PKS-NRPS	NA	14.1	64,805	21,604 – 64,262	42,658	65.8
NRPS	Cupriachelin (11%)	16.1	35,618	77 – 35,618	35,541	99.8
NRPS	NA	21.1	11,311	1 – 11,311	11,311	100
NRPS	NA	22.1	7,911	3,369 – 7,768	4,399	55.6
NRPS	NA	24.1	4,435	1 – 4,435	4,435	100
NRPS	NA	25.1	3,916	1-3,916	3,916	100
NRPS	NA	28.1	1,164	1 – 1,164	1,164	100

Table S5. Alterochromides detected in the EtOAc and 90% aqueous MeOH extracts of JC28 by a short-gradient LC-MS run.

t_R (min)	Observed *m/z*	Remarks
5.24	766.2	Alterochromide A/A'
7.71	844.2	Bromoalterochromide A/A'
8.86	858.2	Bromoalterochromide A/A' + CH_2
9.83	870.2	Bromoalterochromide B/B'
10.27	924.1	Dibromoalterochromide A/A'
10.87	884.3	Bromoalterochromide B/B' + CH_2
12.43	950.2	Dibromoalterochromide B/B'

Chapter 3

Metagenomic analyses of the accessory nidamental gland community provides metagenome-assembled genomes of uncultured members and a source to mine secondary metabolite biosynthetic genes

Introduction

The accessory nidamental gland (ANG) of *Euprymna scolopes* deposits bacteria in the jelly coat (JC) layer of fertilized eggs (69)(72). These bacteria are hypothesized to provide defense against microbial infections (Appendix I), but little is known about their metabolism in the ANG or eggs. Although some isolates from the community have been cultured, some groups have not, and genomic studies have been limited to the *Alphaproteobacteria* (103)(85). Of the thirteen previously published isolates, all are members of the *Alphaproteobacteria*, with ten *Leisingera* spp., two *Ruegeria* spp., and one *Tateyamaria* sp. (103)(85). While the *Leisingera* genus makes up 26.8% of the wild *E. scolopes* ANG community, there are six other genera of *Alphaproteobacteria* present at greater than 0.1% in 16S rRNA amplicon sequencing reads (72) which do not yet have cultured representatives. As well, the second most dominant taxa, the class *Opitutae* of the *Verrucomicrobia* (present at 28.4%), also remain uncultured. When examining just the core community of operational taxonomic units (OTUs) present in wild ANGs (OTUs present in 90% of samples), 29.30% are *Leisingera* spp. and 0.95% are *Opitutus* spp. (72). To gain a better understanding of how this symbiotic community functions in adult hosts and developing eggs, we must first obtain more reference genomes for a greater diversity of the community members.

Metagenomic sequencing of symbiotic communities has provided insights into several symbioses where the majority of community members have not yet been cultured. In cattle, a combination of metagenome binning and Hi-C-based proximity-guided assembly of 43 rumen metagenomes allowed for the assembly of 913 metagenome-assembled genomes (MAGs)(178). These MAGs contained 69,000 proteins predicted to be involved in carbohydrate metabolism of the rumen microbiota, the majority of which only had 60-70% amino acid similarity to proteins

in public databases. Study of these proteins and MAGs will provide future insights into carbohydrate metabolism in rumen microbiomes. Metagenomic sequencing of the Mediterranean sponge, *Aplysina aerophoba*, led to 37 bins from 11 bacterial phyla and revealed an enrichment in genes for defense (toxin-antitoxin systems) and host colonization compared to non-symbiotic relatives (179). As well, the potential for specialized utilization of either host-provided carnitine or sulfated polysaccharides was observed between two symbiont guilds. Analysis of the honey bee gut metagenome revealed that different bacterial members may be specialized to different carbohydrate breakdown pathways and have a high amount of strain level diversity despite a less diverse community at the class level (180).

The goals of this chapter were to sequence ANG metagenomes to bin genomes and predict secondary metabolite biosynthetic genes that may be involved in the production of defensive compounds and understand the metabolism of uncultivated members of the ANG community. A metagenome from one ANG was previously sequenced with 454 pyrosequencing, which yielded 622,987 reads (69). More metagenomic sequencing from four females was desired to capture a more complete picture of the metabolic potential of the ANG bacterial community. Previously sequenced ANG isolate genomes have been analyzed with the antibiotics and secondary metabolite analysis shell (antiSMASH)(181) and possess gene clusters for the production of nonribosomal peptide synthetases (NRPSs), polyketide synthases (PKS), bacteriocins, siderophores, terpenes, ectoines, and homoserine lactones (103)(85). Although these biosynthetic genes have been identified, the structures they produce have not been characterized and the analyzed genomes represent a small percentage of the ANG community. To examine the metabolic potential of the entire community, including members that have not yet been cultured, an antiSMASH analysis was performed on an ANG metagenome co-assembly

of four female's ANGs. These analyses predicted 255 secondary metabolite biosynthetic gene clusters from 18 metabolite classes. Binning of the ANG metagenomes resulted in nine high-quality MAGs, including members of the *Verrucomicrobia*, *Alphaproteobacteria*, and *Gammaproteobacteria*.

Materials and Methods

Sample collection

Adult female Hawaiian bobtail squid, *Euprymna scolopes*, were collected from sand flats in Maunalua Bay, Oahu, Hawaii (21826'3.36"N, 157847'20.78"W) in July 2016 and January 2017. Bobtail squid were shipped back to the University of Connecticut, where they were maintained in aquaria with artificial seawater (Instant Ocean, Blacksburg, VA) and a 12 hr light / 12 hr dark cycle. Four sexually mature females (mantle lengths = 25 – 31 mm, **Table S1**) were sacrificed after one month in the lab following protocols approved by the UConn Institutional Animal Care and Use Committee (IACUC). Dissection tools were soaked in RNase Away (Ambion, Austin, TX) for one hour prior to use. The ANGs were dissected using sterile forceps, cut in half using a sterile razor blade, surface sterilized with 95% ethanol, and rinsed in filter-sterilized squid Ringer's (FSSR, 530 mM NaCl, 25 mM $MgCl_2$, 10 mM $CaCl_2$, 20 mM HEPES, pH 7.5). One half of each ANG sample was placed in RNAlater (Invitrogen, Waltham, MA) and stored at -80°C for future RNA extractions. The other half of each ANG was flash-frozen in liquid nitrogen and stored at -80°C for DNA extractions.

ANG bacterial DNA extraction

Frozen ANG tissue was thawed on ice and homogenized in 0.5 ml FSSR with a ground-glass homogenizer. To pellet squid tissue, homogenate was centrifuged at 100 x g for 5 min., at room temperature. The supernatant containing bacterial cells was removed and centrifuged at 5000 x g for 5 min., at room temperature, to pellet cells. Supernatant was discarded and the bacterial pellet was washed with FSSR until the protein concentration of the supernatant, as measured by absorbance at 280 nm, was sufficiently low (<0.5 mg/ml), totaling six washes.

DNA from the bacterial pellet was extracted using the MasterPure Complete DNA and RNA Purification Kit (Epicentre, San Diego, CA), following the manufacturer's protocol for cell samples with the addition of bead-beating homogenization. Briefly, the bacterial pellet was resuspended in 300 µl of Tissue and Cell Lysis solution with Proteinase K and bead-beat with 0.01 mm and 0.5 mm diameter zirconia/silica beads (BioSpec Products, Bartlesville, OK) for 3 min. using the Mini-Beadbeater-16 (BioSpec Products, Bartlesville, OK). The homogenate was incubated at 65°C for 30 min., shaking at 550 rpm, followed by bead-beating for an additional 3 min., and then incubated for an additional 30 min.. Samples were treated with the kit's RNase for a total of one hour at 37°C. Remaining protein was removed using the kit's MPC protein precipitation reagent and DNA was precipitated using 70% ethanol. DNA was resuspended in TE buffer and stored at -20°C. DNA concentration was measured using the Qubit dsDNA high sensitivity assay kit (Invitrogen, Waltham, MA) and purity was measured as the ratio of absorbance at 260 nm and 280 nm using a Nanodrop spectrophotomer (ThermoFisher Scientific, Waltham, MA). The average DNA concentration was 670 ng/µl ± 901 ng/µl and the average purity was 1.91 ± 0.02 A_{260}/A_{280} (**Table S1**).

Metagenomic Sequencing and Analyses

Metagenomic libraries were prepared for each ANG sample using the TruSeq Nano DNA library prep kit (Illumina, San Diego, CA) following manufacturer's protocols. DNA was sheared for a 550 bp insert size using the S220 focused-ultrasonicator (Covaris, Woburn, MA). Sample PD13 (**Table S1**) was sequenced individually using the MiSeq reagent Nano kit v2 (2x250 cycles) (Illumina, San Diego, CA), yielding 1,569,762 paired end raw reads. All other samples were pooled and sequenced together using the MiSeq reagent kit v2 (2x250 cycles) (Illumina, San Diego, CA), yielding an average of 7.0 ± 0.85 million paired end raw reads (**Table S1**). All samples were sequenced at the Center for Genome Innovation (CGI) at the University of Connecticut.

Raw reads were quality filtered using Trimmomatic v.0.36 (121), removing any reads less than 36 bp, with the following arguments: "LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36". Quality of reads was checked using FastQC v.0.11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmed paired-end reads from each of the four metagenomes were either assembled individually or co-assembled with MEGAHIT v.1.1.2 (182) with a minimum contig length of 1,000 bp and all other default settings. Assembly statistics were reported using QUAST v.4.4 (183) (**Table S2**). Metagenome coverage was determined by mapping reads back to the assemblies using Bowtie2 v.2.3.1 (184) and using the mpileup utility of Samtools v.1.7 (185) to determine the total number of bases covered at a minimum coverage depth of 5x or higher. Breadth of coverage is reported as the percentage of number of bases covered by reads divided by the length of the metagenome assembly (**Table S3**). To screen for secondary metabolite biosynthetic genes, the ANG

metagenome co-assembly was analyzed with the command-line version of antiSMASH v.3.0 (124). Taxonomy of the assembled contigs was assigned using Kraken v.1.0 (186).

Individual sample metagenome assemblies were manually binned in the Anvi'o analysis platform v5.4 (187) based on contig taxonomy, read coverage, and percent genome completion and redundancy values. The metagenome co-assembly was binned using CONCOCT (188) in Anvi'o. Taxonomy of a bin was assigned based on the taxonomy of the majority of contigs. Metagenome assembled genomes (MAGs) were assessed in bins based on percent genome completion greater than 90% and percent contamination less than 10%, assessed using Anvi'o, and in accordance with the Minimum Information about a Metagenome-Assembled Genome (MIMAG) recommendations (189). Percent genome completion is based on the number of single-copy housekeeping genes present from the list provided by (190). The number of single-copy housekeeping genes that were present in a bin more than once was used to calculate the percent redundancy and is used as an indicator that more than one genome may be present in a bin. MAGs were annotated with RAST (123) and screened for secondary metabolite biosynthetic genes using antiSMASH v5 (181)(191).

Phylogenetic Analysis of *Verrucomicrobia* MAGs

Pairwise average nucleotide identities (ANI) of the *Verrucomicrobia* MAGs were calculated using JSpeciesWS with BLAST+ v.2.2.29 (192) to determine if MAGs represented similar species. The two PD13 *Verrucomicrobia* MAGs were placed in a phylogenetic tree with 39 *Verrucomicrobia* genomes available in Genbank. The tree was reconstructed from a concatenated alignment of amino acid alignments of five housekeeping genes (*dnaA*, *dnaG*, *dnaN*, *fmt*, *rpsB*) present in all genomes. Individual alignments were generated using MUSCLE

(126) with default parameters and concatenated in Geneious v.10 (Biomatters Inc., Newark, NJ). The best evolutionary model, LG+F+I+G4, was selected using ModelFinder (127) in the IQtree webserver (128). A maximum likelihood tree was constructed using IQtree v.1.6.5 (129) with 1000 bootstraps (130). The tree was visualized using the Interactive Tree of Life (iTOL) v.4 webserver (https://itol.embl.de/) (131). *Kiritimatiella glycovorans* L21-Fru-AB was selected as an outgroup due to sub-division five's proposed reassignment to a novel sister phylum (193)(194).

Metabolic model of *Verrucomicrobia* PD13 Bin 1

Genes for different metabolic pathways were categorized using the SEED designations assigned in RAST (123). Some genes that were unassigned by RAST were manually assigned a category based on homology (percent amino acid similarity > 30%) to known proteins in the non-redundant NCBI database from a DIAMOND v.0.9.9 BLASTp analysis (195). Carbohydrate active enzymes were identified using the dbCAN meta server (http://bcb.unl.edu/dbCAN2/).

***Verrucomicrobia* PD13 Bin 1 multi-copy gene trees**

Due to the high number of copies of sulfatases, arylsulfatases, and alpha-L-fucosidases located in the *Verrucomicrobia* PD13 Bin 1 genome, and the nature of this draft being a community-assembled genome, protein trees were created for each of these genes to compare their evolutionary relation. For each gene, amino acid sequences from other *Verrucomicrobia* genomes or distantly related organisms were retrieved from UniProt and aligned using MUSCLE (126) with default parameters. ModelFinder (127) in the IQtree webserver (128) was used to select the best evolutionary model for each alignment. The model Blosum62+F+R5 was selected

for the sulfatase alignment, Blosum62+F+R9 for the arylsulfatase alignment, and LG+F+R4 for the alpha-L-fucosidase alignment. Maximum likelihood trees were constructed using IQtree v.1.6.5 (129) with 1000 bootstraps (130). The trees were visualized using the Interactive Tree of Life (iTOL) v.4 webserver (https://itol.embl.de/) (131).

Results

Metagenome assemblies

MEGAHIT assemblies of the individual sample metagenomes ranged in size from 33.6 Mbp up to 83.7 Mbp, while the co-assembly of all ANG samples had a total length of 228.5 Mbp (**Table S2**). Although sample PD13's individual assembly had the smallest size (33.6 Mbp), it yielded the longest contig of 3,381,902 bp. The assemblies had a percent GC content that ranged from 55.4% – 57.5%. The individual assemblies had a high percent breadth of sequencing coverage, ranging from 85.8% up to 92% of the assembly being covered at 5x depth or higher (**Table S3**). In the co-assembly, the breadth of coverage from each sample ranged from 11.3% up to 44.87%, reflecting the varying number of reads from each sample used as input.

Taxonomy was assigned to 35,592 of a total 63,844 contigs (55.75%) in the ANG metagenome co-assembly. Of these assigned contigs, 99.91% were bacterial, 0.04% archaeal, and 0.02% viral (**Fig. 1A**). Kraken uses the rule of lowest common ancestor to assign taxonomy, and thus 0.03% of contigs were assigned to "root," meaning they were classified into multiple domains of life and could not be more specifically assigned. Focusing on just the bacterial assignments, contigs for all classes that were previously observed in *E. scolopes* ANGs (69)(72) were detected here. The majority of contigs (88.4%) were assigned to *Proteobacteria*, followed by *Bacteroidetes* (6.2%), *Verrucomicrobia* (1.8%), and *Actinobacteria* (1.3%) (**Fig. 1A**). The

other 25 phlya were at relative abundances below 1% (**Fig 1A**). Of the *Proteobacteria*, the most abundant classes were the *Alphaproteobacteria* (85.6%) and *Gammaproteobacteria* (10.4%) (**Fig. 1B**). As was previously seen by 16S rRNA gene sequencing and 454 shotgun metagenomics (69)(72), the community is dominated by the orders *Rhodobacterales* (63% of all contigs) and *Rhizobiales* (7.3% of all contigs) (**Fig. 1C**).

Metagenome-assembled genomes (MAGs)

Binning of the ANG metagenome co-assembly with CONCOCT yielded a total of 61 bins, with genome completion rates ranging from 0 – 98.6% (**Fig. 2, Table S4**). Only four bins (bins 17, 35, 40, 41) had a percent completion above 90%, and of those, three bins had a percent redundancy below 10%: *Verrucomicrobia* bin 34, *Vibrio* bin 40, and *Alphaproteobacteria* bin 41 (**Tables 1, S4**). Overall 47.5% of bins were assigned to *Alphaproteobacteria*, 14.8% *Gammaproteobacteria*, 8.2% *Flavobacteriia*, 6.6% *Verrucomicrobia*, and 21.3% of bins could not be assigned to a taxa (**Table S4**).

Since some co-assembly bins had high coverage from only one sample, it appeared that some MAGs would be better assembled from individual metagenome assemblies rather than the co-assembly. Because there is variability in the ANG community at the genus level (72), and likely at the strain level between females, and metagenome sequencing captures only a portion of the communities genes, combining samples that are not deeply re-sequenced in the assembly could introduce more strain variation that can make it difficult to pull out individual, complete MAGs (178). From the individual assemblies, six high quality MAGs were recovered, including 3 *Verrucomicrobia* spp., 1 *Erythrobacter* sp., 1 *Mesorhizobium* sp., and 1 *Vibrio* sp. (**Table 1**). The 16S rRNA gene was recovered from five MAGs, and a nucleotide BLAST comparison

showed that *Mesorhizobium* E7 Bin 6 had 97% similarity to *Mesorhizobium tamadayense*, *Verrucomicrobia* co-assembly Bin 40 was 89% similar to *Cephaloticoccus capnophilus*, and *Opitutae* E7 Bin 10, *Verrucomicrobia* PD13 Bin 1, and *Verrucomicrobia* PD13 Bin 2 were all 87-88% similar to *Coraliomargarita akajimensis*. The 23S rRNA gene was recovered from *Opitutae* E7 Bin 10 and *Verrucomicrobia* PD13 Bin 1, and both were 84% similar to an uncultured verrucomicrobium. Only the 5S rRNA gene was recovered from *Vibrio* E8 Bin 5 and *Vibrio* co-assembly Bin 40, and both were 100% similar to *Vibrio parahaemolyticus* MAVP-Q. There was still difficulty binning many of the contigs to the point where a bin could be identified below the class level, however some bins that did not have MAGs with high percent genome completion could be identified as *Leisingera, Ruegeria, Rhizobium,* and *Muricauda* genera (**Table S5, Figures S1-S4**).

Secondary metabolite biosynthetic genes detected in the ANG metagenome co-assembly and MAGs

The ANG metagenome co-assembly was predicted to contain 255 secondary metabolite biosynthetic gene clusters (**Fig. 3a, Table 2, Appendix III**). Of these clusters, 25.1% were predicted to make homoserine lactones (HSLs), 17.3% bacteriocins, 15.7% terpenes, 10.2% nonribosomal peptide synthetases (NRPSs), 8.6% polyketide synthases (PKSs), 4.3% ectoines, 3.5% aryl polyenes, 2.7% siderophores, 1.2% thiopeptides, 0.4% cyanobactins, 0.4% acyl amino acids, and 0.4% nucleosides (**Table 2, Appendix III**). Based on antiSMASH's clusterblast analysis, 164 clusters were most similar to clusters from *Alphaproteobacteria*, 19 clusters to *Flavobacteriia*, 15 to *Gammaproteobacteria*, 4 to *Bacilli*, 3 to *Actinobacteria*, 3 to *Betaproteobacteria*, 2 to *Cyanobacteria*, 2 to *Verrucomicrobia*, 1 to *Clostridia*, 1 to

Deltaproteobacteria, and 41 clusters had no significant match. Comparing this to the Kraken taxonomy of the contigs, 146 clusters were found on *Alphaproteobacteria* contigs, 13 on *Gammaproteobacteria,* 10 *Flavobacteriia,* 6 *Actinobacteria,* 5 *Betaproteobacteria,* 2 *Bacilli,* 2 unclassified bacteria, 2 unclassified *Proteobacteria,* 1 *Chitinophagia,* 1 *Opitutae,* 1 *Bacteroidetes,* and 66 clusters were on contigs that did not receive a taxonomy identification.

Nine of the terpene clusters had 100% of the genes share similarity to carotenoids from *Rhodobacterales* strains, including one match to a previously isolated ANG symbiont, *Tateyamaria* sp. ANG-S1 (103). Five terpene clusters had 28% of the genes share similarity with carotenoids from *Flavobacteriales* strains. One other terpene cluster had 75% of the genes share similarity with an astaxanthin dideoxyglycoside from a *Sphingomonadales* strain. Two aryl polyene clusters had 11% and 19% of genes share similarity to flexirubins from *Flavobacteriales* strains. Two other aryl polyene clusters had 15% and 85% of genes share similarity to aryl polyene carboxylic acids from *Vibrionales* and *Xanthomonadles* strains. One "other" cluster had 20% of the genes share similarity to a siderophore, Scabichelin, from a Rhodobacterales strain.

Thirteen genomes from ANG symbiont isolates have previously been sequenced (196)(103)(85) and 69 of the predicted metagenome clusters are homologous to clusters in twelve of these genomes (**Appendix III**). Homologous clusters included 24 HSLs, 15 bacteriocins, 6 siderophores, 6 "others", 5 ectoines, 5 T1-PKSs, 3 terpenes, 3 thiopeptides, and 1 NRPS. These clusters were homologous to *Leisingera* spp. ANG1, ANG-DT, ANG-M1, ANG-M6, ANG-M7, ANG-S, ANG-S3, ANG-S5, ANG-Vp; *Ruegeria* spp. ANG-R, ANG-S4; and *Tateyamaria* sp. ANG-S1. The greatest number of clusters was homologous to *Ruegeria* sp. ANG-R (5 HSLs, 3 ectoines, 2 bacteriocins, 2 terpenes, 1 thiopeptide), followed by *Ruegeria* sp. ANG-S4 (7 HSLs, 2 bacteriocins, 1 NRPS, 1 thiopeptide, 1 other).

The predicted secondary metabolite gene clusters in the MAGs had similar results to the overall metagenome co-assembly (**Fig. 3b, Table S6**). *Alphaproteobacteria* co-assembly bin 41 had 11 predicted clusters (3 HSLs, 3 bacteriocins, 1 bacteriocin-lanthipeptide, 1 terpene, 1 ectoine 1 betalactone, 1 NRPS-T1PKS). *Erythrobacter* E7 bin 3 had 2 clusters (1 bacteriocin, 1 terpene). *Mesorhizobium* E7 bin 6 had 4 clusters (2 HSLs, 1 NRPS, 1 terpene). The *Opitutae*/ *Verrucomicrobia* MAGs had 2 to 4 predicted clusters, including NRPS, NRPS-like, and "other" type clusters. The *Vibrio* MAGs had 4 and 6 clusters, including NRPSs, betalactones, aryl polyenes, and bacteriocins.

Verrucomicrobia PD13 Bin 1 Metabolic Model

The *Verrucomicrobia* are the second most dominant community member in the ANG, but due to difficulty getting this isolate into culture, very little is known about its metabolism in *E. scolopes*. ANI analysis revealed that *Verrucomicrobia* PD13 bin 1 and *Verrucomicrobia* co-assembly bin 35 are highly similar (ANI > 99%) and are likely the same species (**Table S7**). *Verrucomicrobia* PD13 bin 2 and *Opitutae* E7 bin 10 shared 85% ANI similarity, but are below the species level cutoff (ANI > 95%) (**Table S7**). All other MAG pairwise comparisons were less similar (70-71% ANI), indicating that the *Verrucomicrobia* MAGs may represent three different species. Of the *Verrucomicrobia* MAGs, PD13 bin 1 was the most contiguous (4 contigs), had a high percent genome completion (96.4%), low redundancy of single-copy house-keeping genes (1.4%), and had both the 16S and 23S rRNA genes. For these reasons, this MAG was chosen for further genomic analyses.

Verrucomicrobia PD13 bin 1 had 4,095 coding sequences, with 2,090 of those predicted to be hypothetical proteins. 57 tRNAs were present in the genome for all 20 proteinogenic amino

acids. Of the key cell division proteins, genes were detected for *ftsZ* (Z-ring), *ftsA* (assists in Z-ring assembly), *ftsH* (zinc metallopeptidase), *ftsW* (peptidoglycan polymerase), *ftsQ* (may link cytoplasmic/periplasmic division proteins), *ftsX* (ABC transporter involved in septal ring formation), and *ftsY* (involved in targeting/insertion of nascent membrane proteins). Cell division genes that were not detected in the assembly include the activator of septal peptidoglycan synthesis (*ftsN*), the DNA translocase (*ftsK*), and the membrane subcomplex proteins, *ftsBL* (197). The chromosomal replication initiator protein, *dnaA*, and partitioning proteins, *smc* and *parAB*, were present. The rod-shape determining proteins, *mreBC*, were detected along with peptidoglycan synthetases, binding proteins, and glycosyltransferases.

Motility and chemotaxis genes were predicted in the *Verrucomicrobia* PD13 bin 1 MAG. Of the 21 core flagellar biosynthesis genes (198), 13 genes were present, including *fliC* (main filament protein), *fliD* (filament cap), *flgE* (hook), *flgG* (distal rod), *flgB* (proximal rod), *flhAB* / *fliIPQR* (export apparatus), and *motAB* (motor). In addition to the core genes, several other flagellar genes were present: *flgD* (hook-capping protein), *flgH* (L-ring), *flgI* (P-ring), and *flbT* (biosynthesis repressor). The chemotaxis genes, *cheBDRX*, were present, suggesting that this strain can respond to chemoattractants. In addition to flagellar genes, four copies each of the twitching motility proteins, *pilBT*, and one copy of *pilC*, were present. These genes are involved in assembly of the type IV pilus, which can be utilized in twitching motility, biofilm formation, surface attachment, or natural transformation.

The PD13 bin 1 MAG had 139 transport-related genes, making up 3.4% of all genes (**Fig. 4**). Of the transport genes, 58 were ABC transporters, including 13 gene copies for components of the antimicrobial peptide transport system, five efflux transporters, oligopeptide transporters (*oppABCDF*), and the periplasmic substrate-binding component (*dppD*) of the dipeptide

transporters. Other ABC transport genes included transporters for ferric iron, molybdenum, nitrate, sulfonate, nickel, bicarbonate, phospholipids, vitamin B12 (*btuF*), and ribose (*rbsABC*). The other transport-related genes included gliding motility-associated transport genes, *gldAF*; permeases for sulfate, fucose, acetate (*actP*), phosphate (*pstAC*/*phoU*); transporters for sialic acid, ammonium, biopolymers (*exbD*/*tolR*), N-acetylglucosamine (*nagX*), ferrous iron, maltose (*malT*), xyloside (*xynT*), lysophospholipid (*lplT*), galactose, mannose, and cations; and symporters for L-rhamnose, glucose, choline, sulfate, and sodium/solutes. Exporters were present for lysine (*lysE*/*yggA*), magnesium and cobalt (*corC*), potassium (*kefA*), proteins (*secA*), and RND multidrug efflux systems. Other transport systems included TonB-dependent transporters, TRAP-type transporters (for C4-dicarboxylate), and the twin-arginine translocation proteins, *tatAC*.

An analysis using the carbohydrate-active enzyme (CAZy) database assigned 302 genes as members of 68 glycoside hydrolase (GH) families, making up 7.4% of all genes in the genome (**Fig. 4, Table S8**). The most abundant GH families included the alpha-N-acetylgalactosaminidases (GH109), alpha-L-fucosidases (GH29), beta-galactosidases (GH2), alpha-1,3-L-neogarooligosaccharide hydrolases (GH117), beta-glucanase precursors (GH16), sialidases (GH33), endo-1,4-beta-xylanase (GH10), and beta-agarases (GH50). With the exception of GH2, 36% to 81% of the genes in these GH families had signal peptides, suggesting they are secreted to utilize environmental carbon sources. The enrichment in these GH families suggests a preference for the utilization of mucin glycoproteins and plant, in particular algal, polysaccharides. N-acetylgalactosamine is one of the core glycans in mucins, and alpha-N-acetylgalactosaminidases can cleave these glycans from the serine/threonine amino acids of the mucin backbone (199). Fucosidases of the GH29 and GH95 families can cleave fucose from *O*-

glycosidic chains linked to galactose or N-acetyl-glucosamine in mucins (200). The neuraminidases and sialidases of the GH33 family can cleave sialic acid from the non-reducing ends of mucin oligosaccharides, usually constituting the first step of mucin degradation (200). The PD13 Bin 1 MAG also possesses sialic acid transporters and sialic acid 9-O-acetylesterases, which can transport sialic acid into the cell and remove the O-acetyl ester groups. The MAG also has genes to import cleaved fucose via fucose permeases and begin transforming this carbon source via a L-fucose-isomerase, L-fucose kinase, and L-fucose mutarotase. Although mucin utilization via glycosyl hydrolases (in particular GH95 genes) has been best described in *Bifidobacterium* (201), it has also been associated with *Verrucomicrobia*, such as the human gut symbiont, *Akkermansia muciniphila* (202)(203).

Sulfate can also be present on mucins, and although not categorized by CAZy because they do not act on glycosidic bonds, sulfatases can play an important role in mucin utilization. The PD13 Bin 1 MAG had 255 sulfatases (**Fig. 4**), 184 of which are arylsulfatases, which can act on sulfated mucin glycans. The assimilative sulfate reduction genes were also present, such as sulfate permeases (brings sulfate into the cells), sulfate adenylyltransferases (converts sulfate to adenylylsulfate [APS]), adenylylsulfate kinase (converts APS to 3'-phosphoadenylylsulfate [PAPS]), phosphoadenylyl-sulfate reductase (converts PAPS to sulfite), and sulfite reductases (converts sulfite to hydrogen sulfide). The hydrogen sulfide produced in this pathway can be utilized to produce cysteine, as a cysteine synthase gene was detected.

Due to the high number of sulfatases, arylsulfatases, and alpha-L-fucosidases in the PD13 Bin 1 genome, single gene trees were constructed based on amino acid alignments with other bacterial and non-bacterial copies of these genes (**Fig. S5-S7**). Although many of these gene copies are related to other copies in the genome, and similar genes in other *Verrucomicrobia*

genomes, no two genes appear to be duplicates of each other. Due to the metagenomic source of this MAG, however, it cannot be ruled out that this high copy number is due to presence of contigs from multiple strains in this bin.

Phylogenetic placement of *Verrucomicrobia* PD13 Bin 1 and assignment as a member of the *Puniceicoccales*

Placement of *Verrucomicrobia* MAGs PD13 Bin 1 and Bin 2 in a phylogenetic tree with 39 other *Verrucomicrobia* genomes, based on five housekeeping genes, revealed that these strains did form a branch together within the *Puniceicoccales* (**Fig. 5**). This order is a member of the *Opitutae* class and sub-division 4 of the *Verrucomicrobia* phylum. Although the 16S rRNA genes for both of these MAGs were most similar to *Coraliomargarita akajimensis*, the *Coraliomargarita* genomes formed a separate sister branch within this sub-division. The closest relatives to the PD13 MAGs were *Puniceicoccales* bacterium Verruco_01 and '*Candidatus* Moanabacter tarae TARA' strain B10000123. The *Puniceicoccales* Verruco_01 genome was assembled from a pool of 6 metagenomes of anaerobic cellulose-degrading digesters (BioSample: SAMN03837578). The genome size is 2.5 Mb with 63% GC content. The '*Ca*. M. tarae' genome was assembled from the Tara Oceans Consortium's metagenomic sequencing of epipelagic water surrounding the Marquesas Islands in the Pacific Ocean (204). This strain has a genome size of 2.6 Mb, 46% GC content, and the 16S rRNA gene was similar to marine sediment *Verrucomicrobia.* Other *Puniceicoccaceae* MAGs have been assembled from Russian soda lake metagenomes (BioProject: PRJNA434545) and Mediterranean Sea metagenomes (BioProject: PRJNA352798). The closest cultured relative to the *E. scolopes* MAGs is *Coraliomargarita sinensis* WN38, which was isolated from a marine solar saltern in Weihai,

China (205). This strain is a non-motile obligate anaerobe with a genome size of 3.2 Mb and 54.7 mol% GC content.

The *E. scolopes* PD13 MAGs fall within a branch that is dominated by genomes from environmental sources (**Fig. 5**). The majority of the *Coraliomargarita* strains and "*Ca.* M. tarae" strain are from marine seawater or sediment samples. Since the ANG symbionts are horizontally acquired, these *Verrucomicrobia* strains would also need to be adapted to a free-living lifestyle, fitting with their placement amongst other free-living marine *Opitutae* strains. Overall, there was no strict clustering of environmental versus host-associated strains in the tree (**Fig. 5**), although there is clustering among strains isolated from the same host, such as the *Opitutaeae* bacteria TAV1, TAV2, TAV3, and TAV5 from termite hindguts, and *Cephaloticoccus primus* CAG34 and *Cephaloticoccus capnophilus* CV41 from *Cephalotes* ants.

Discussion

Metagenome binning and Verrucomicrobia MAGs

From binning ANG metagenomes, nine high-quality MAGs were assembled, eight of which represent genomes from ANG bacterial taxa that have not yet been sequenced. Although the majority of *Alphaproteobacteria* bins could not be resolved to MAGs, one *Erythrobacter* bin and one *Mesorhizobium* bin were assembled. These genomes are still highly fragmented, with 4,424 *Erythrobacter* contigs and 97 *Mesorhizobium* contigs, but they are of particular interest because they represent two new genomes from the ANG community. The *Verrucomicrobia* MAGs are also of high interest since no members of this phylum have been cultured from the ANG, yet it is the second most abundant member of the community. Although also fragmented

to varying degrees (4 up to 483 contigs), these genomes add to the growing availability of all *Verrucomicrobia* genomes—a group that has been traditionally difficult to culture in labs (193).

The *Verrucomicrobia* are ubiquitous in terrestrial and aquatic environments (206) but little is known about their ecological roles. Of 517 water samples collected from all oceans, *Verrucomicrobia* were present in 98% of samples and average 2% of the water column and 1.4% of sediment communities (206). They can be found in a wide range of conditions, from freshwater to marine, strict aerobes to anaerobes. They have a wide variety of lifestyles, from intracellular symbionts to free-living, motile or non-motile, and vary in genome size. In aquatic environments, they are more abundant in shallow coastal and brackish waters, and may be more frequently associated with particles (based on filtering methods before sequencing) (206).

The *Verrucomicrobia* were first described as a new division, and *Verrucomicrobiae* a new class, from the *Prosethecobacter* and *Verrucomicrobium* genera in 1997 (207). Currently, the *Verrucomicrobia* belong to the *Planctomyces*/ *Verrucomicrobia*/ *Chlamydiae* (PVC) superphylum, which also includes the *Lentisphaerae* and *Poribacteria* (104)(208). This division was initially divided into five sub-divisions by Hugenholtz et al. (209), but two more sub-divisions were later introduced: subdivision 6 (represented by one 16S rRNA gene sequence)(210) and sub-division 7 (represented by the *Methylacidiphilum* spp. isolated from volcanic pools)(211). Sub-divisions 1 (*Verrucomicrobiae*) and 4 (*Opitutae*) dominate marine environments, while subdivisions 2 (*Spartobacteria*) and 3 (*Opitutae*) are more frequent in soil communities (206). Sub-division 5 includes one genus, *Kiritimatiella*, and has recently been proposed as a novel phylum, *Kiritimatiellaeota* (193). The *Puniceicoccales* order, *Puniceicoccus* genus, and *Opitutae* class, to which the PD13 Bin 1 and Bin 2 MAGs were assigned, were first described by Hugenholtz et al. (212). The type strain, *Puniceicoccus vermicola* IMCC1545^{T}, is a

non-motile, facultative anaerobe, isolated from the gut of a marine clamworm, *Periserrula leucophyrna,* from a tidal flat in the Yellow Sea. Although no genome has been sequenced yet, the GC content was chemically analyzed to be 52.1 mol% (212). This strain was found to produce carotenoid pigments with absorbance peaks at 320 nm and 480 nm. No carotenoid biosynthesis genes were detected in the *E. scolopes* MAGs. Unlike the type strain, the PD13 Bin 1 MAG does have the majority of flagellar biosynthesis genes and may be motile.

The larger genome sizes of some *Verrucomicrobia* have been associated with the metabolic potential to utilize many different carbon sources. The PD13 Bin 1 and Bin 2 MAGs have larger genomes than most *Opitutae* (5.1 Mb and 4.16 Mb respectively, compared to the 2.5 Mb genome of *Puniceicoccales* bacterium Verruco_01 or 3.2 Mb of *Coraliomargarita sinensis*). However, other *Puniceicoccales* MAGs from freshwater reservoir metagenomes also had genome sizes ranging from 1.3 Mb up to 5.69 Mb (213). Many *Verrucomicrobia* genomes have high numbers of glycoside hydrolases, and it has been suggested they play a role in polysaccharide degradation in freshwater ecosystems (194). The GH95 (mannosyl-oligosaccharide alpha-1,2-mannosidase) and GH2 (beta-galactosidase) family genes that were abundant in the PD13 Bin 1 MAG were also very abundant in freshwater lake *Verrucomicrobia* MAGs (194). GH family genes made up 0.4% - 4.9% of these freshwater MAGs, and sulfatases also made up a significant portion of the genome for some, ranging from less than 0.5% up to 3% of all genes. In the PD13 Bin 1 MAG, GH genes made up the highest percentage of all genes (7.4%), with sulfatases following close behind at 6.3% of the genome (**Fig. 4**). A higher number of sulfatases can be expected for the *E. scolopes* MAGs due to their presence in a marine environment where more sulfate would be present, which is seen in other marine or hyper saline *Verrucomicrobia* (193). RNA-sequencing of the ANG environment will reveal which of the

carbohydrate metabolism pathways are expressed in the host and may provide clues for the best lab culture conditions.

Secondary metabolite genes in the ANG metagenome

The antiSMASH analysis of the ANG co-assembly predicted 255 secondary metabolite biosynthetic gene clusters from 18 classes. These included, in order of abundance, homoserine lactones, bacteriocins, terpenes, NRPSs, PKSs, ectoines, aryl polyenes, siderophores, thiopeptides, and a cyanobactin and nucleoside clusters. While seven of these classes have been detected in *Alphaproteobacteria* ANG isolate genomes previously (103)(85), they were identified in a wider range of bacteria in the ANG metagenome. Sixty-nine of the metagenome clusters were homologous to clusters in all of the ANG isolate genomes. Outside of the *Alphaproteoabacteria*, 19 clusters had similarity to clusters in *Flavobaccteriia,* 15 clusters were similar to *Gammaproteobacteria*, 3 to *Betaproteobacteria*, and 2 to *Verrucomicrobia*. Other clusters had similarity to bacteria not previously reported in ANG communities (*Cyanobacteria*, *Actinobacteria, Bacilli,* and *Deltaproteobacteria*), however it is possible that these clusters have not yet been described in any of the bacteria that make up the ANG community. It is also possible that some clusters may be horizontally acquired. For example, a cyanobacteria tunicate symbiont, *Synechocystis trididemni,* is believed to have horizontally acquired biosynthesis genes for the cyclic peptide, didemnin, as its gene cluster was reported in the *Alphaproteobacteria*, *Tistrella mobilis and Tistrella bauzanensis* (31). Future analyses could determine if transposable elements flank these clusters as evidence of horizontal transmission.

Of the known clusters that were predicted, nine terpene clusters had 100% similarity to carotenoids in *Alphaproteobacteria* and *Flavobacteriia*. These clusters were expected because

the orange-red coloration of the ANG tissue was previously reported to originate from bacterial carotenoids (74). Only six other clusters had matches to other known clusters, including four aryl polyenes, which were found to be the most widespread and abundant class of secondary metabolites in 1,154 publicly available bacterial genomes (214). It is not surprising that the majority of clusters did not have known structures, as that same study of publicly available genomes also could not identify a known structure for the majority of their 33,351 putative biosynthetic gene clusters (214). These unknowns make the ANG bacterial community a promising source of novel secondary metabolites. Coupling genome mining with analyses of chemical extracts will hopefully characterize these unidentified structures.

Additional metagenomic sequencing with long-read sequencing could help disambiguate some of the predicted clusters. 55.7% of the predicted clusters were either at the edge of a contig or encompassed the entire contig. Due to the repetitive domains that characterize a number of these biosynthetic enzymes, such as in the NRPSs/PKSs, short reads can cause assemblers to misassemble or fragment these clusters. The addition of longer reads could increase the number of predicted clusters or produce complete biosynthetic gene clusters, as has been demonstrated in *Pseudonocardia* genome sequencing (215).

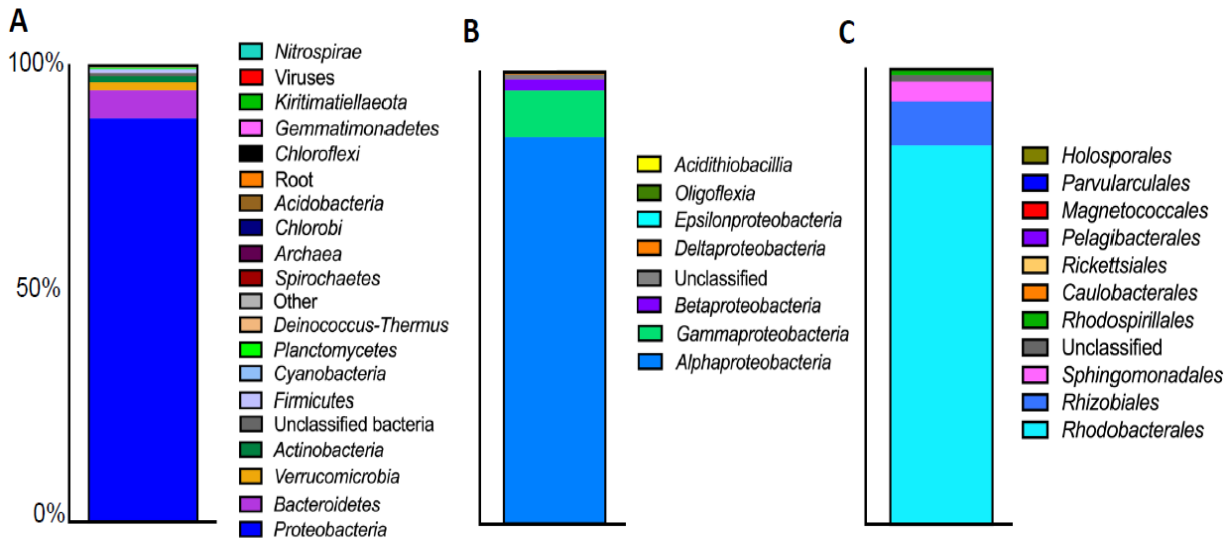


Figure 1. Taxonomy of ANG metagenome co-assembly contigs. (A) Relative abundances of bacteria at the phylum level. Viruses remain at the domain level. Root indicates contigs that matched to multiple domains. (B) Relative abundance of classes in the *Proteobacteria* phylum. (C) Relative abundance of orders in the *Alphaproteobacteria* class.

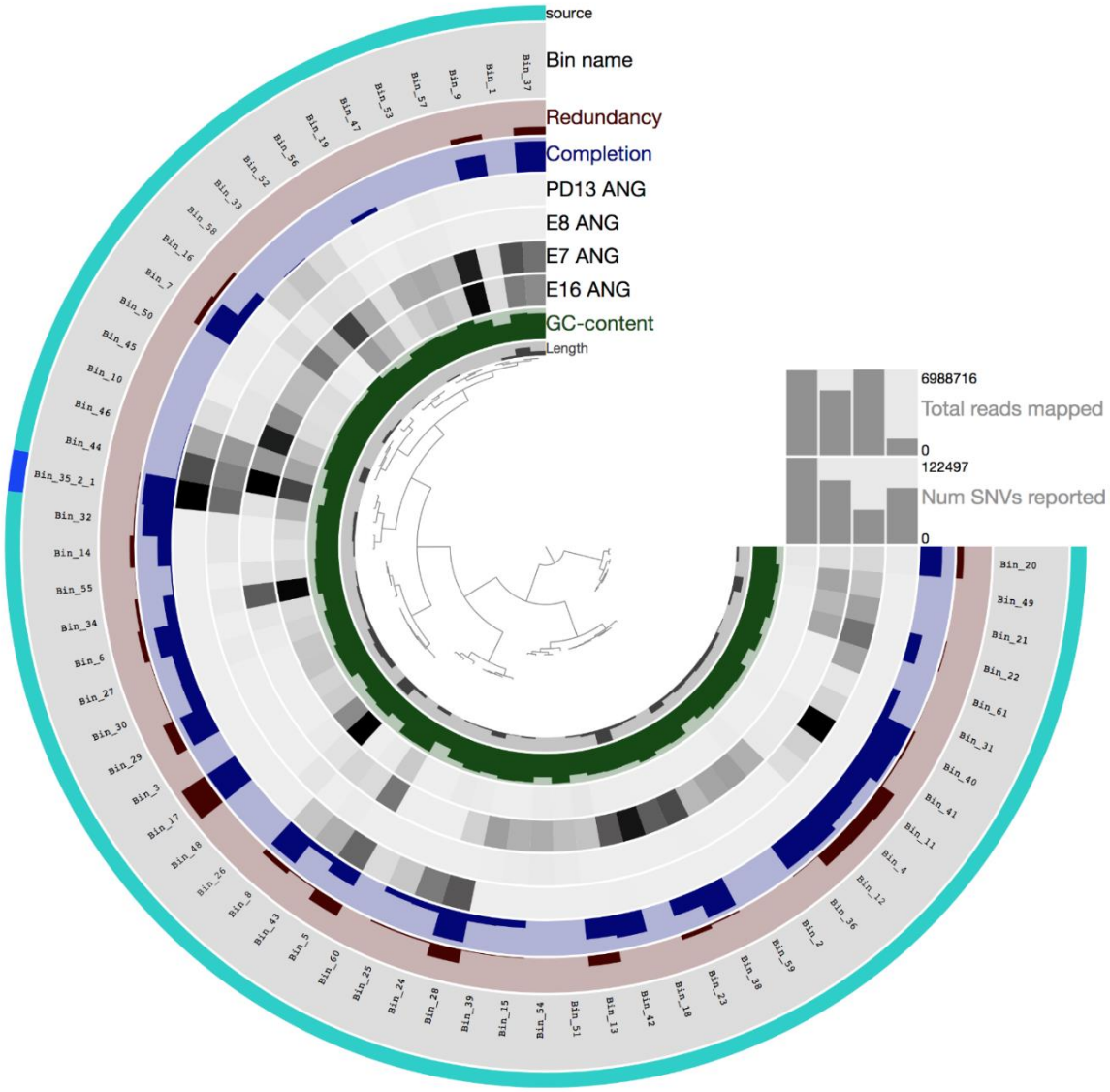


Figure 2. ANG metagenome co-assembly CONCOCT bins. Each leaf in the dendrogram represents a bin. Working outward, length is the total bin length in Mbp, GC-content is the average percent of GC bps in the bin, gray E16/E7/E8/PD13 bars represent the coverage of each sample's reads mapped to the bin's contigs, completion is the percent of single-copy housekeeping genes present in the bin, redundancy is the percent of single-copy housekeeping genes that were present in a bin more than once, bin name, source: all contigs were automatically binned with CONCOCT (teal), or automatically binned then manually curated in Anvi'o (dark blue).

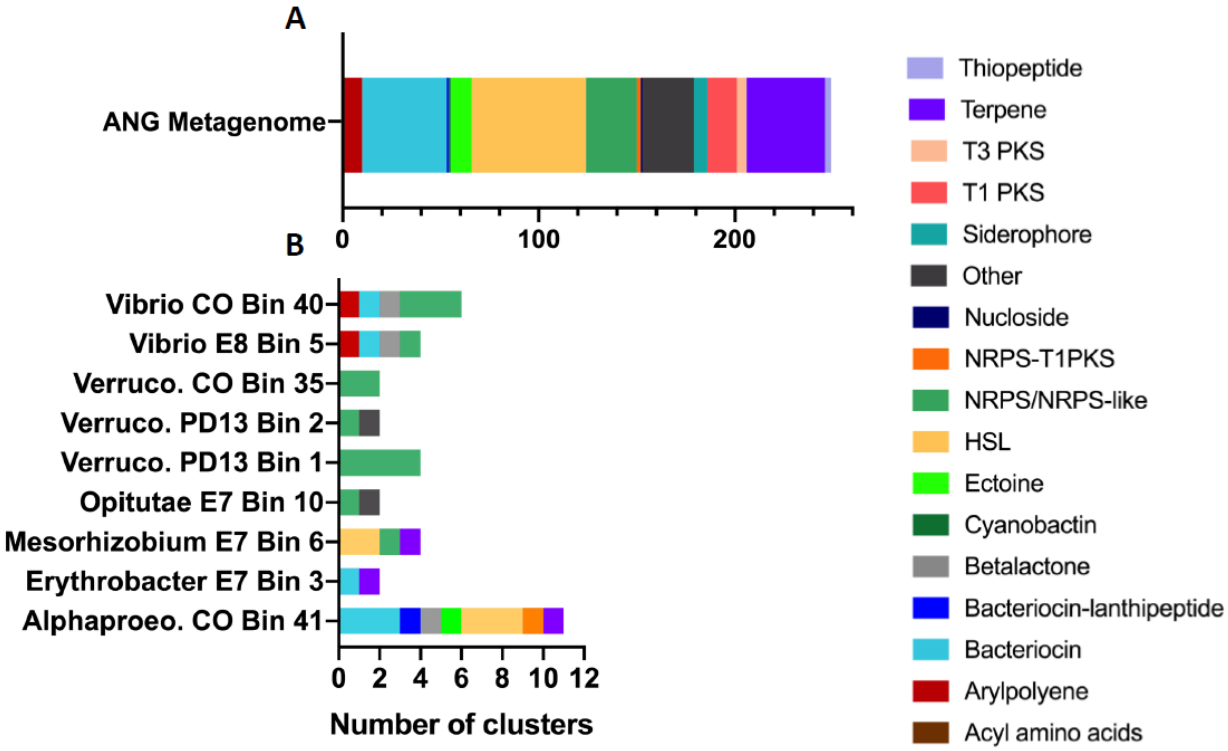


Figure 3. Predicted secondary metabolite biosynthetic genes in the ANG metagenome co-assembly and MAGs. (A) Secondary metabolite biosynthetic gene clusters predicted by antiSMASH v.3.0 in the ANG metagenome co-assembly. (B) Clusters predicted by antiSMASH v.5.0 in the *Alphaproteobacteria*, *Erythrobacter*, *Mesorhizobium*, *Opitutae*, *Verrucomicrobia*, and *Vibrio* MAGs. Abbreviations: CO = co-assembly, *Verruco.* = *Verrucomicrobia*

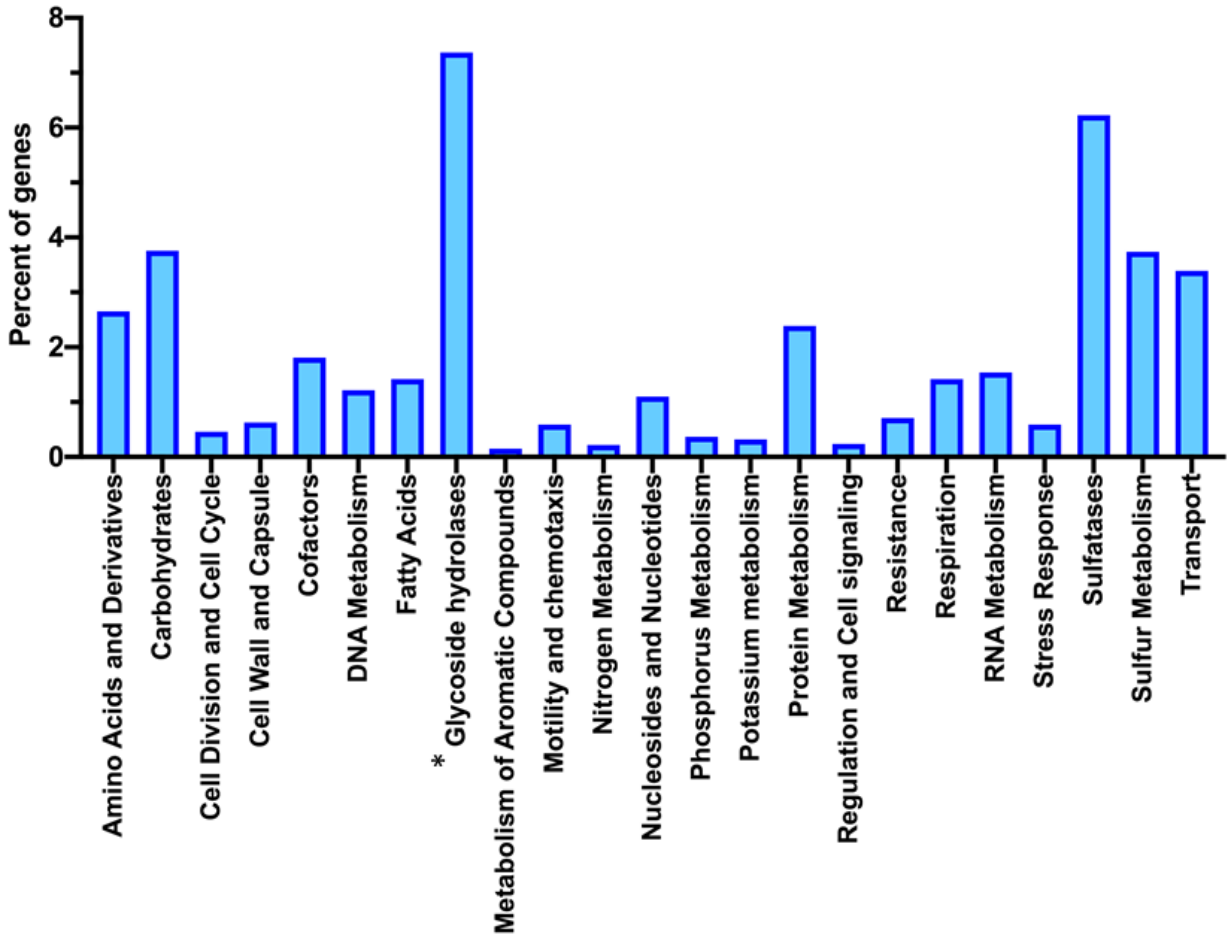


Figure 4. Percent of genes in different SEED categories for *Verrucomicrobia* PD13 Bin 1 MAG. Number of genes in each SEED category is predicted by RAST. Number of glycoside hydrolases (indicated by the *) is predicted by dbCAN. Number of sulfatases was manually counted based on RAST annotations.

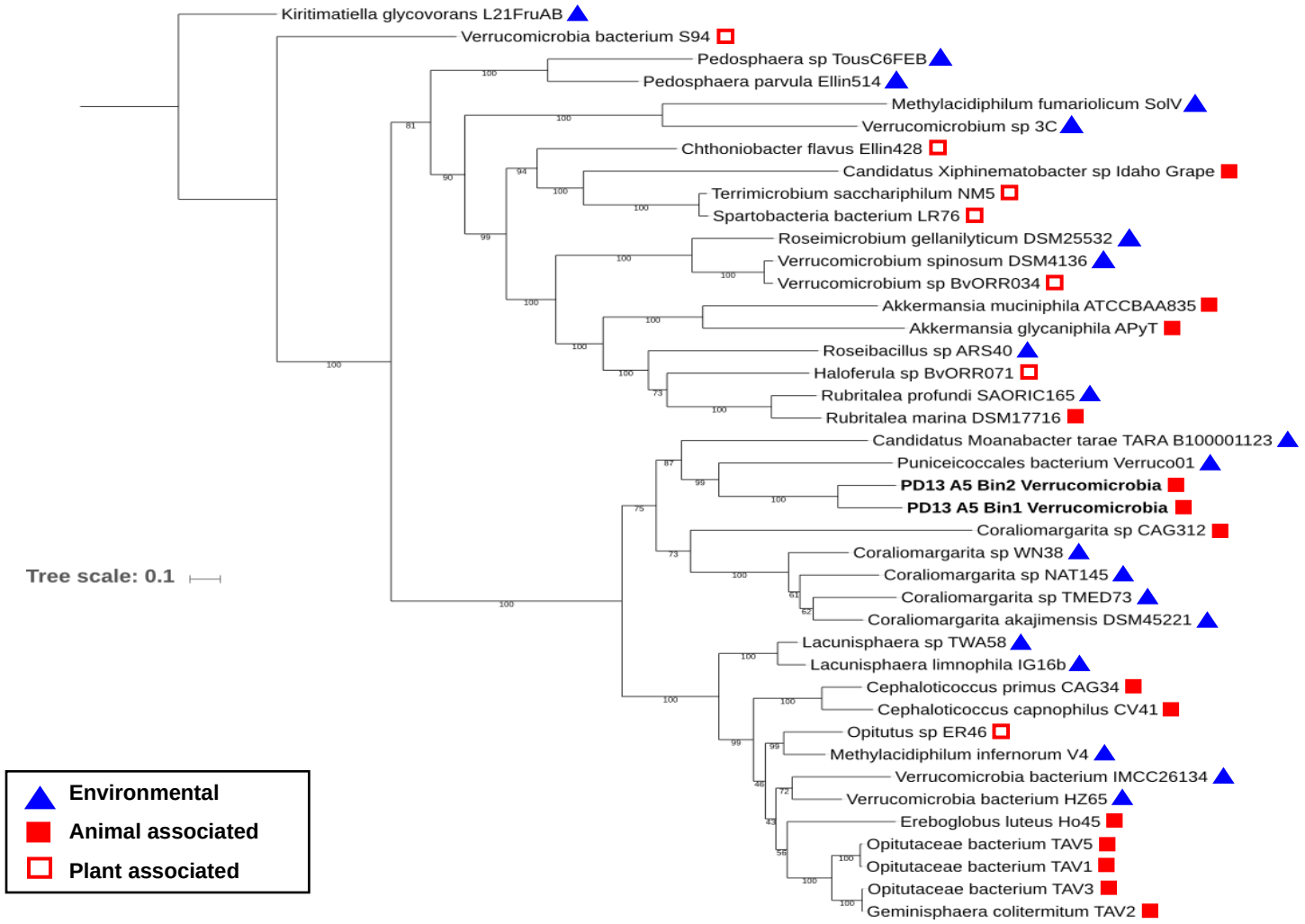


Figure 5. *Verrucomicrobia* maximum-likelihood phylogenetic tree. Phylogenetic analysis was performed with 5 single-copy, housekeeping genes. PD13 Bin 1/Bin 2 MAGs (bold) fall within the *Puniceicoccales.* Isolate/MAG isolation source indicated by blue triangle (environmental), filled red square (animal associated), or outlined red square (plant associated).

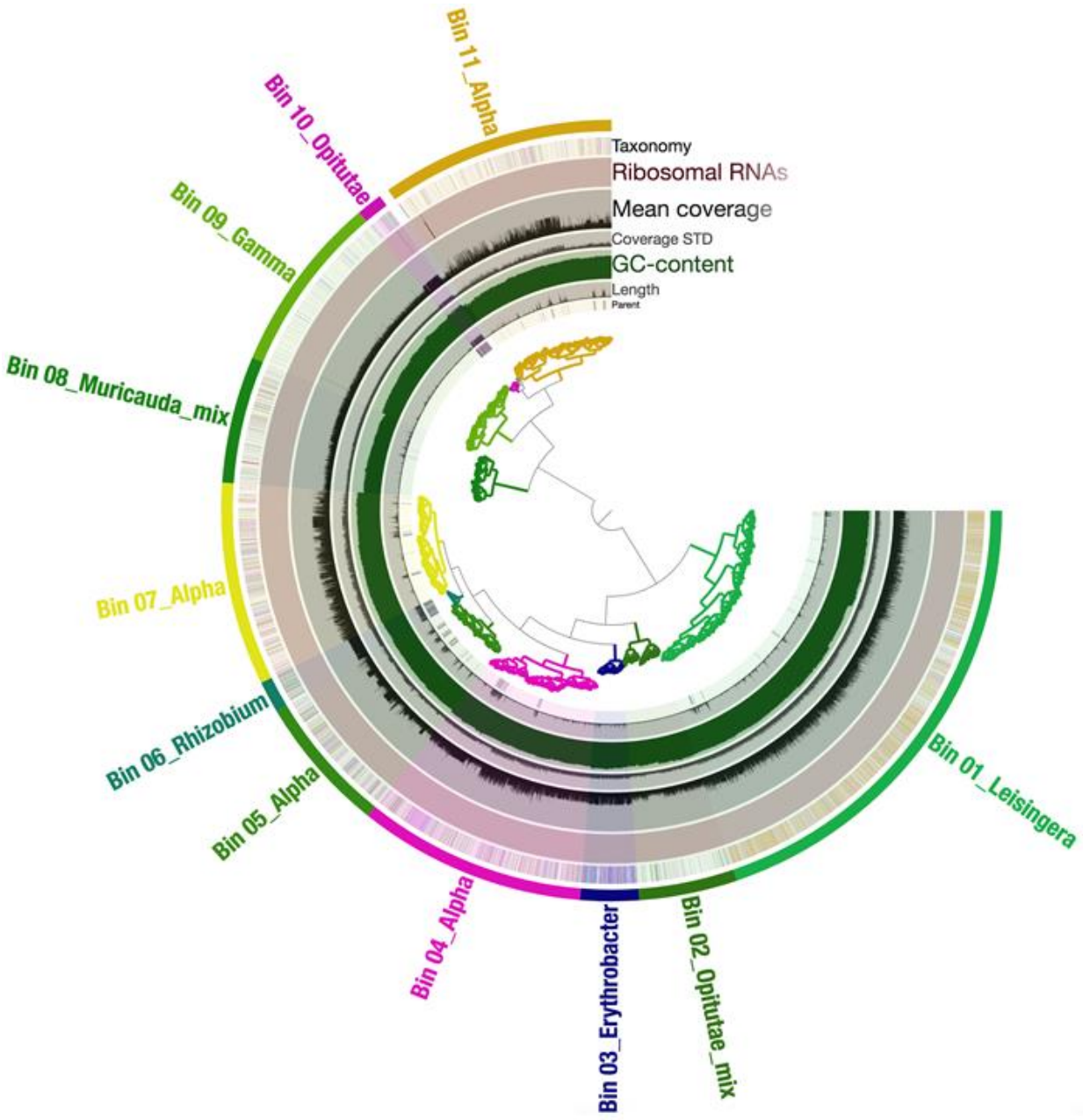
Table 1. Metagenome-assembled genome (MAG) statistics

Taxonomy	Assembly	Bin	Total size (Mb)	Number of contigs	Percent completion	Percent redundancy	GC content (%)
Alphaproteobacteria	Coassembly	41	5.89	93	92.8	4.3	58.54
Erythrobacter	E7 individual	3	2.78	4424	97.1	6.5	64.09
Mesorhizobium	E7 individual	6	4.82	97	99.3	1.4	56.34
Opitutae	E7 individual	10	4	45	97.8	0.7	56.34
Verrucomicrobia	PD13 individual	1	5.1	4	96.4	1.4	52.5
Verrucomicrobia	PD13 individual	2	4.16	131	98.6	0	57.2
Verrucomicrobia	Coassembly	35	4.6	483	91.4	2.2	52.93
Vibrio	E8 individual	5	4.91	151	100	5	44.77
Vibrio	Coassembly	40	5.2	223	98.6	6.5	44.49

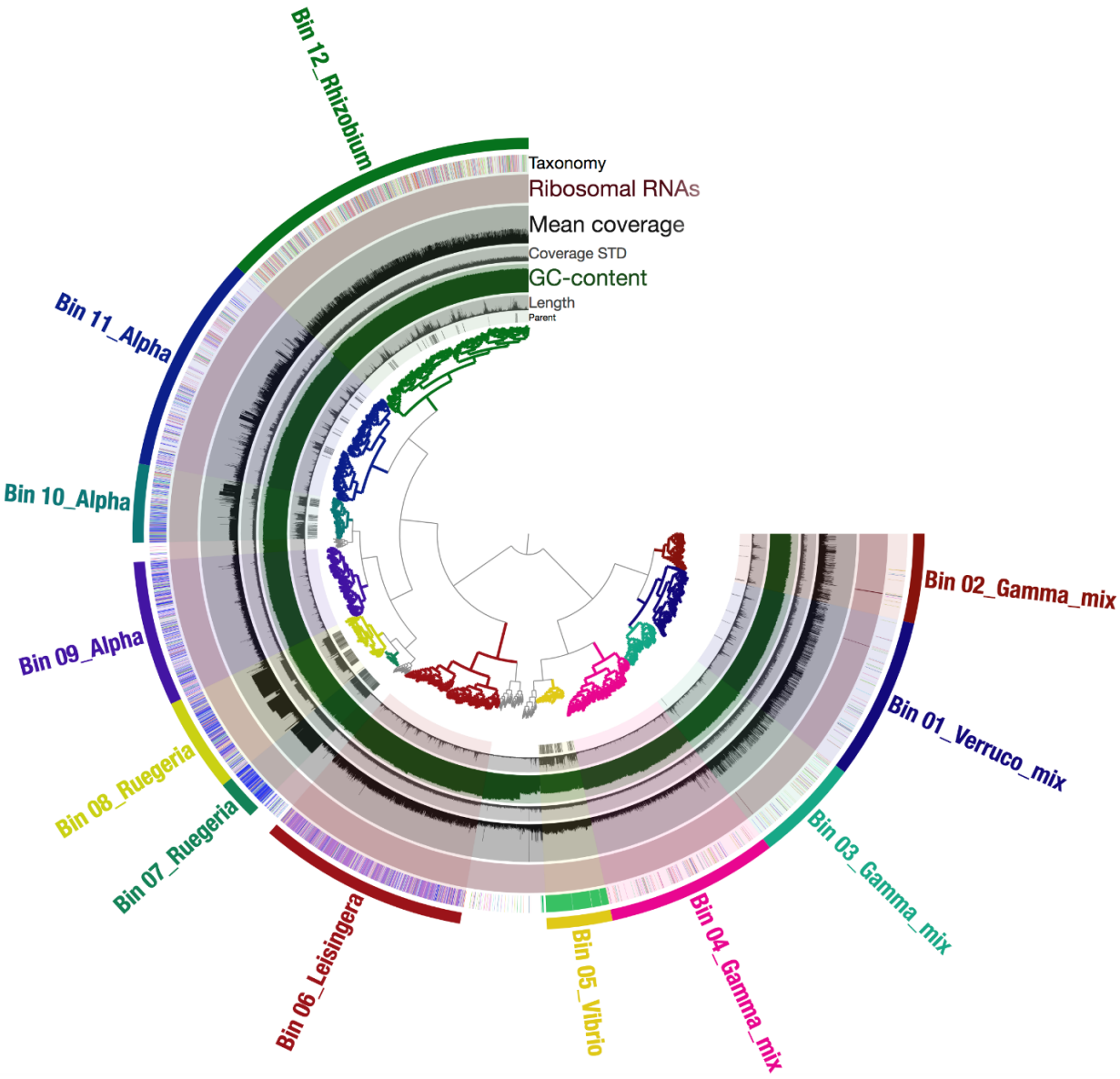
Table 2. Predicted secondary metabolite biosynthetic gene clusters in the ANG metagenome co-assembly

Cluster type	Number of clusters	Class of top homologous hit
Acyl amino acids	1	*Alphaproteobacteria*
Arylpolyene	3	*Flavobacteriia*
	2	*Alphaproteobacteria*
	2	*Gammaproteobacteria*
	1	unclassified (Phylum: *Verrucomicrobia*)
Arylpolyene-resorcinol	1	*Flavobacteriia*
Bacteriocin	30	*Alphaproteobacteria*
	7	*Gammaproteobacteria*
	3	No significant ClusterBlast hits found
	2	*Betaproteobacteria*
	1	*Flavobacteriia*
Bacteriocin-lantipeptide	1	*Alphaproteobacteria*
Cyanobactin	1	No significant ClusterBlast hits found
Ectoine	11	*Alphaproteobacteria*
Homoserine lactone	53	*Alphaproteobacteria*
	9	No significant ClusterBlast hits found
	1	*Betaproteobacteria*
	1	*Gammaproteobacteria*
NRPS	8	No significant ClusterBlast hits found

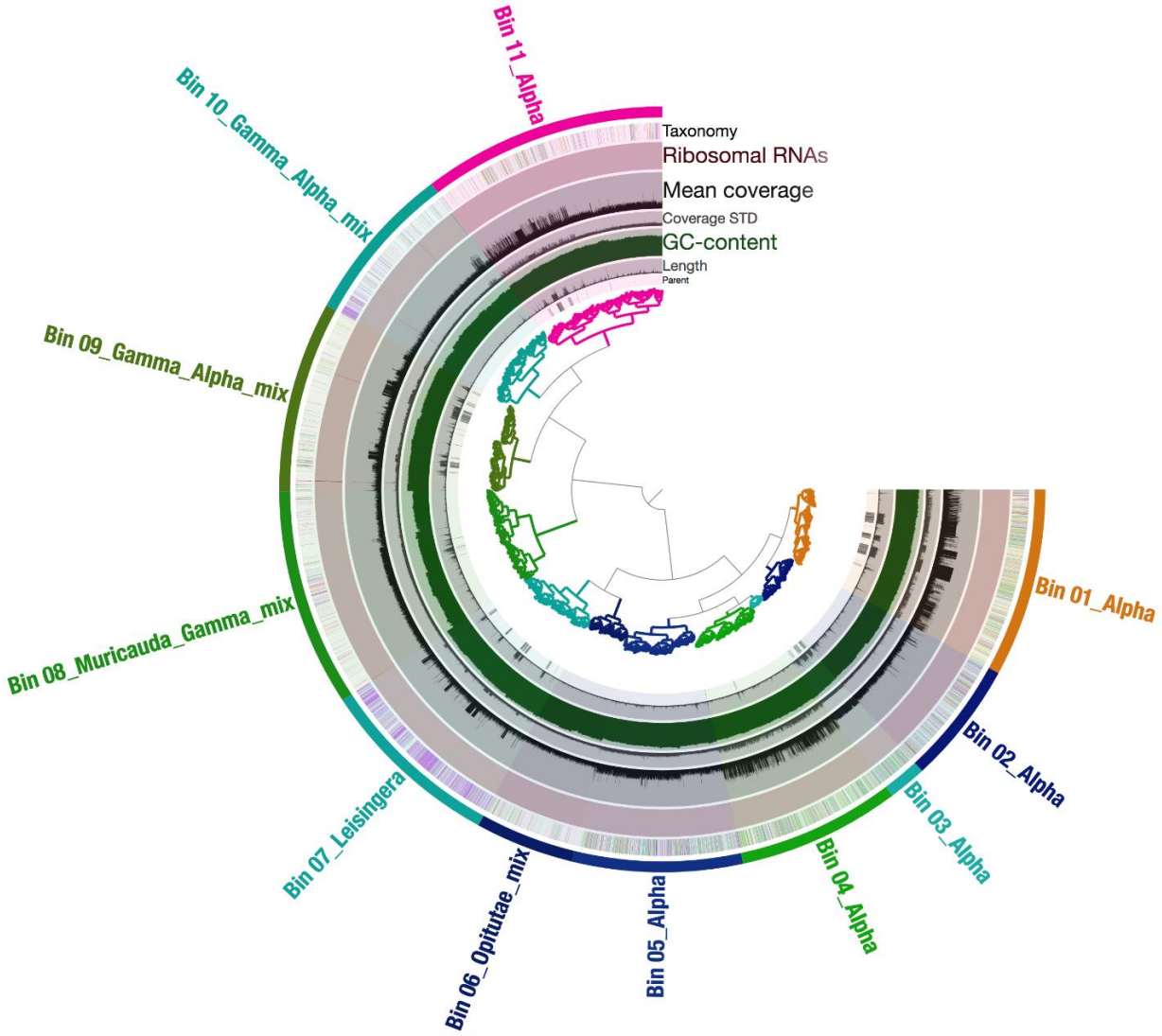
	4	*Alphaproteobacteria*
	4	*Bacilli*
	4	*Gammaproteobacteria*
	2	*Flavobacteriia*
	1	*Actinobacteria*
	1	*Cyanobacteria*
	1	*Deltaproteobacteria*
	1	*Opitutae*
Nucleoside	1	*Actinobacteria*
Other	14	*Alphaproteobacteria*
	8	No significant ClusterBlast hits found
	1	*Actinobacteria*
	1	*Clostridia*
	1	*Cyanobacteria*
	1	*Gammaproteobacteria*
Siderophore	6	*Alphaproteobacteria*
	1	No significant ClusterBlast hits found
T1PKS	8	No significant ClusterBlast hits found
	7	Alphaproteobacteria
T1PKS-NRPS	1	Alphaproteobacteria
	1	No significant ClusterBlast hits found
T3PKS	2	Flavobacteriia
	1	Alphaproteobacteria
	1	No significant ClusterBlast hits found
T3PKS-arylpolyene	1	Flavobacteriia
Terpene	30	Alphaproteobacteria
	9	Flavobacteriia
	1	No significant ClusterBlast hits found
Thiopeptide	3	Alphaproteobacteria
Total clusters:	255	



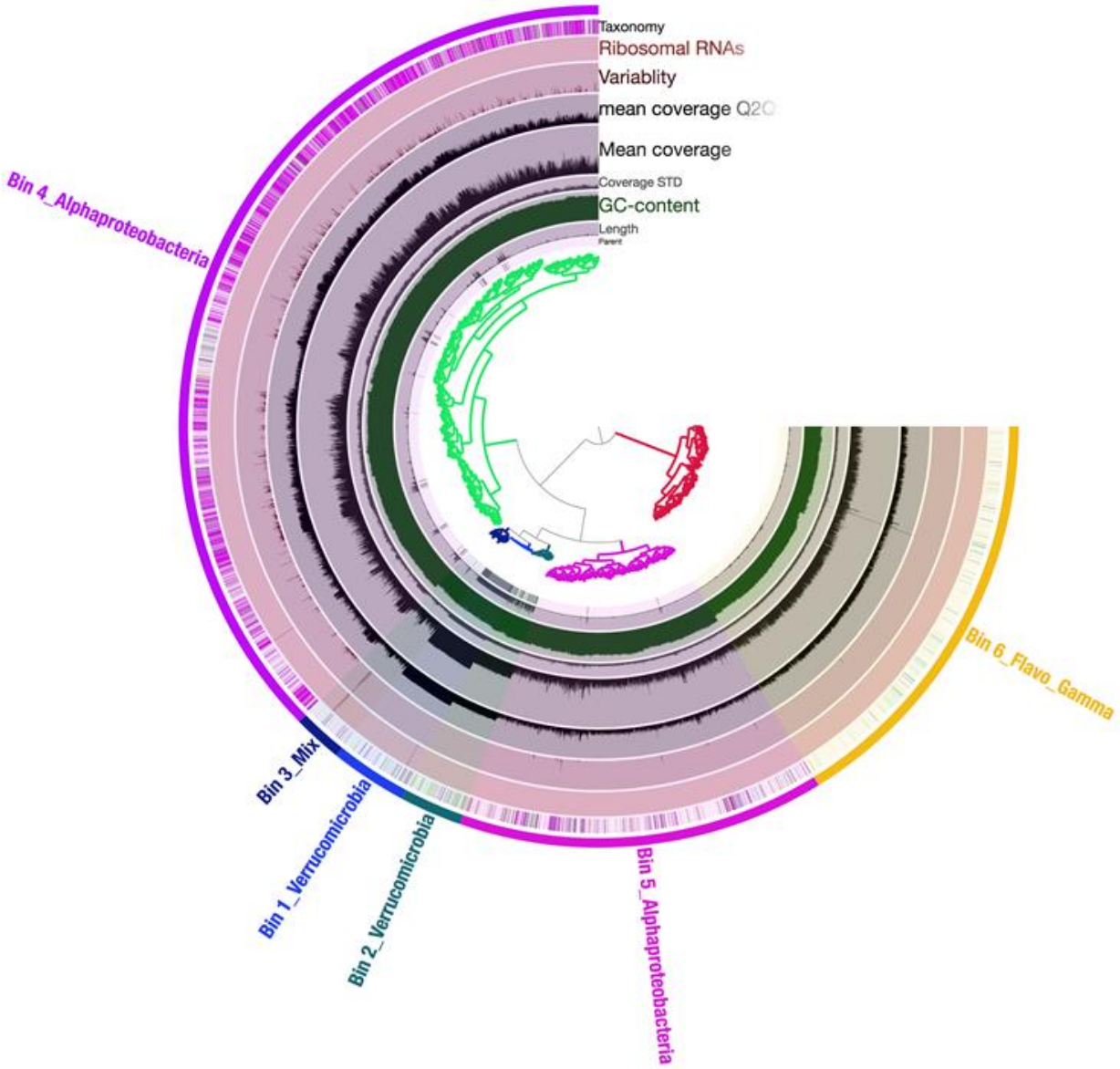
Supplemental Figure S1. Manual binning of ANG metagenome sample E7 in Anvi'o.



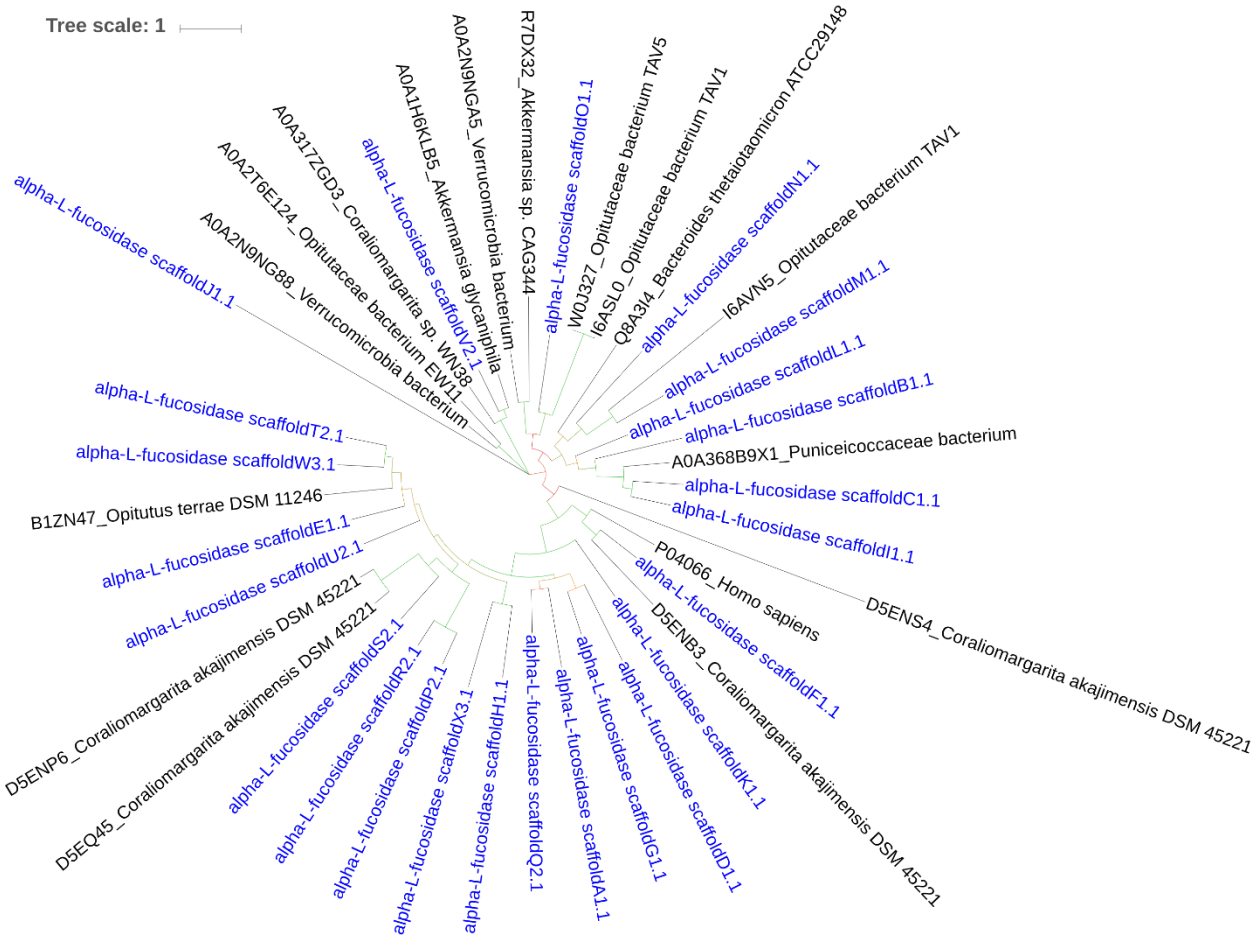
Supplemental Figure S2. Manual binning of ANG metagenome sample E8 in Anvi'o.



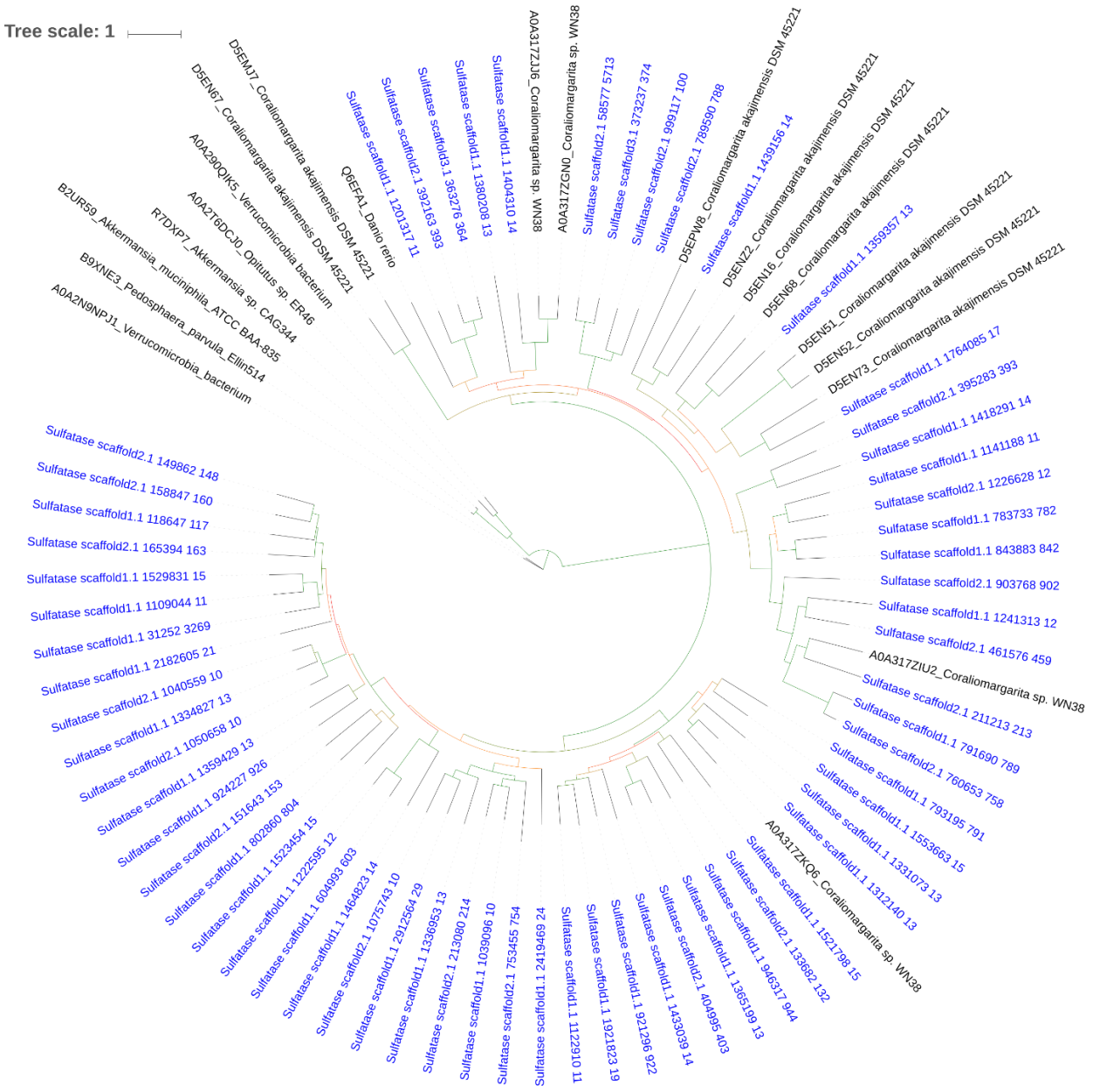
Supplemental Figure S3. Manual binning of ANG metagenome sample E16 in Anvi'o.



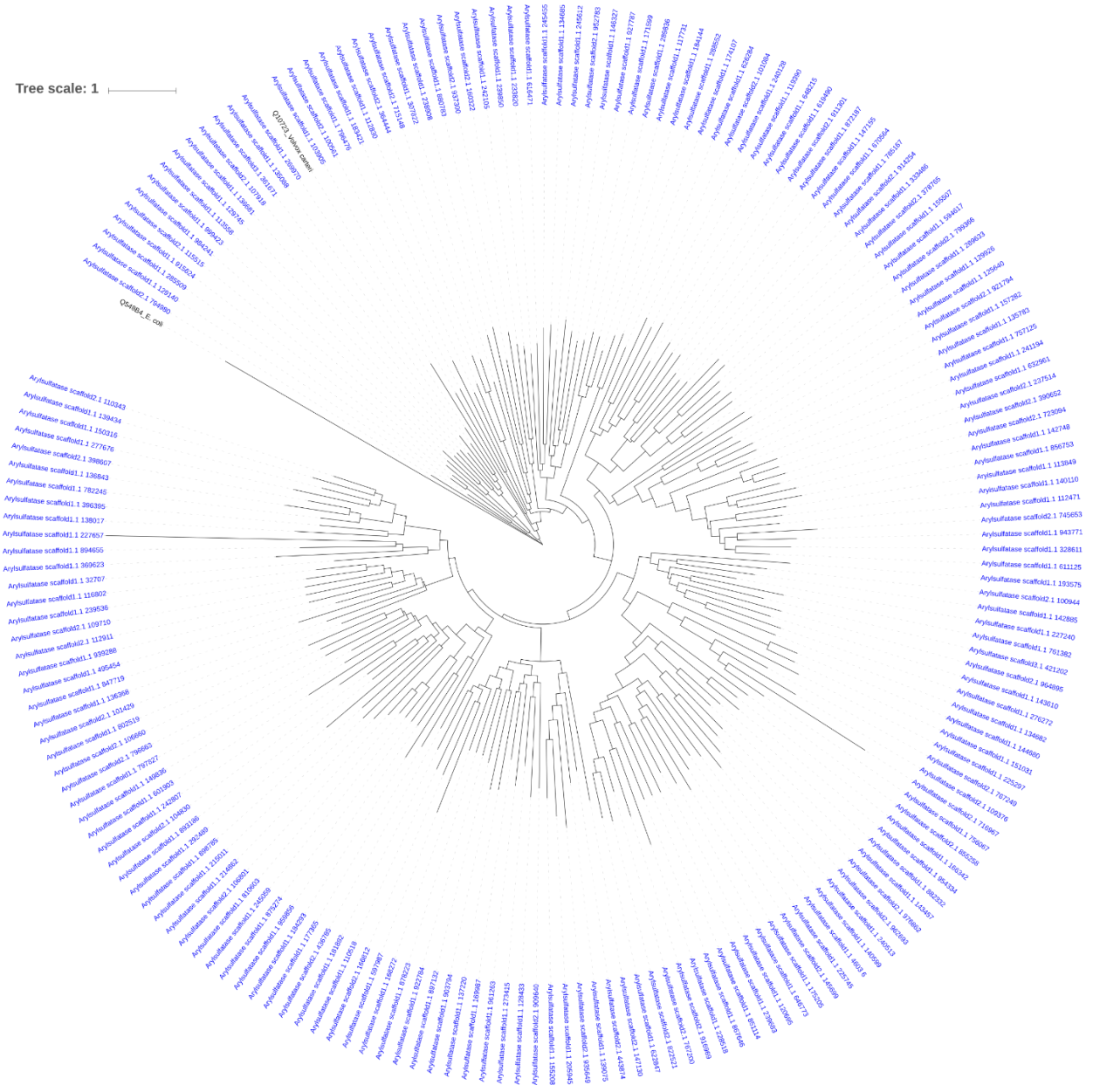
Supplemental Figure S4. Manual binning of ANG metagenome sample PD13 in Anvi'o.



Supplemental Figure S5. Maximum-likelihood protein tree of alpha-L-fucosidase genes in *Verrucomicrobia* PD13 Bin 1 MAG.



Supplemental Figure S6. Maximum-likelihood protein tree of sulfatase genes in *Verrucomicrobia* PD13 Bin 1 MAG.



Supplemental Figure S7. Maximum-likelihood protein tree of arylsulfatase genes in *Verrucomicrobia* PD13 Bin 1 MAG.

Table S1. ANG metagenome samples

Female	Mantle Length (mm)	DNA concentration (ng/µl)	DNA purity (A260/A280)	Number of raw reads (bp)	Number of quality filtered reads
PD13	31	2,020	1.88	1,569,762	1,552,904
E7	26	274	1.92	6,069,860	5,945,198
E8	25	200	1.92	7,337,052	7,212,224
E16	25	185	1.9	7,673,484	7,444,656

Table S2. ANG metagenome assembly statistics

Assembly	Number of contigs	Total length (bp)	Largest contig (bp)	N50	%GC
PD13	22780	33,555,507	3,381,902	7,349	57.17
E7	12619	83,670,993	439,779	9,718	57.46
E8	16589	73,898,804	1,456,288	29,109	56
E16	11947	77,962,384	555,804	8,990	55.41
Coassembly	63844	228,541,819	712,356	5,679	55.81

Table S3. ANG metagenome assembly coverage

Reads	Reference assembly	Overall Bowtie2 alignment rate (%)	Total number of bases covered at ≥ 5X coverage	Length of reference	Breadth of coverage (%)
PD13	PD13 individual	91.29	30,856,129	33,555,507	91.96
E7	E7 individual	76.9	109,159,940	127,233,802	85.79
E8	E8 individual	92.39	64,983,039	73,898,804	87.94
E16	E16 individual	84.92	69,832,694	77,962,384	89.57
PD13	Coassembly	87.2	25,828,646	228,541,819	11.30
E7	Coassembly	87.86	102,556,581	228,541,819	44.87
E8	Coassembly	95.93	63,887,167	228,541,819	27.95
E16	Coassembly	91.03	86,609,212	228,541,819	37.90

Table S4. ANG metagenome co-assembly CONCOCT bins

Bin	Total size (Mbp)	Number of contigs	Percent completion	Percent redundancy	GC content (%)	Taxonomy
1	12.5	6852	0	0	59.78	*Alphaproteobacteria*
2	9.68	6375	0	0	58.1	*Alphaproteobacteria*
3	9.84	5626	0	0	58.94	*Alphaproteobacteria*
4	4.88	3171	71.9	41.7	57.68	*Alphaproteobacteria*
5	5.11	2712	45.3	37.4	63.45	*Leisingera*
6	3.53	2335	74.8	16.5	46.26	*Vibrio*
7	4.63	2148	85.6	16.5	63.44	*Opitutus*
8	4.02	2121	85.6	15.8	42.39	*Muricauda*
9	3.28	2106	62.6	15.8	41.72	*Flavobacteriia*
10	10.75	1988	0	0	62.13	*Leisingera*
11	6.59	1672	67.6	47.5	56.19	*Ruegeria*
12	5.86	1485	87.1	40.3	48.53	*Gammaproteobacteria*
13	2.11	1408	58.3	29.5	50.19	*Alphaproteobacteria*
14	2.03	1370	39.6	12.9	50.32	*Gammaproteobacteria*
15	2.11	1366	20.8	1.4	63.91	*Opitutaceae*
16	2.4	1325	28.8	10.1	61.95	*Leisingera*
17	8.24	1270	97.1	79.9	42.75	*Muricauda*
18	19.94	1038	0	0	57.25	*Alphaproteobacteria*
19	2.01	1172	14.4	0.7	63.43	*Leisingera*
20	3.17	1140	64	19.4	52.76	None
21	13.29	969	0	0	57.24	*Alphaproteobacteria*
22	3.26	1053	36	2.9	61.61	*Ruegeria*
23	4.24	1044	48.2	13.7	56.33	*Ruegeria*
24	1.75	1022	18.7	4.3	63.06	*Leisingera*
25	1.38	973	16.5	4.3	55.93	*Phaeobacter*
26	17.24	577	0	0	54.39	*Phaeobacter*
27	3.34	875	42.4	1.4	53.19	None

28	7.67	823	85.6	40.3	53.19	*Leisingera*
29	7.67	809	85.6	32.4	58.97	*Ruegeria*
30	3.67	809	43.9	2.2	53.25	*Gammaproteobacteria*
31	1.02	740	24.5	0.7	59.38	*Verrucomicrobia*
32	4.09	700	84.2	2.9	49.6	*Gammaproteobacteria*
33	1.07	661	2.4	0	60.04	*Leisingera*
34	1.6	638	33.1	9.4	55.83	*Ruegeria*
35	4.6	483	91.4	2.2	52.93	*Verrucomicrobia*
36	3.87	573	83.5	2.2	39.77	*Muricauda*
37	6.22	443	88.5	23	57.43	*Alphaproteobacteria*
38	2.61	439	89.2	4.3	63.97	*Erythrobacter*
39	0.653	417	22.3	2.2	64.74	*Alterythrobacter*
40	5.2	223	98.6	6.5	44.49	*Vibrio*
41	5.89	93	92.8	4.3	58.54	*Alphaproteobacteria*
42	3.94	100	66.9	0	56.28	None
43	0.202	145	18.7	3.6	64.25	None
44	0.152	60	2.5	0	41.03	*Vibrio*
45	0.097	60	0	0	60.81	*Alphaproteobacteria*
46	0.087	58	1.8	0	52.4	None
47	0.087	52	0	0	61.38	*Alphaproteobacteria*
48	0.092	29	0	0	51.23	*Gammaproteobacteria*
49	0.031	25	0	0	61.02	*Alphaproteobacteria*
50	0.033	23	0	0	63.78	None
51	0.04	22	0	0	65.97	None
52	0.031	21	0	0	51.77	None
53	0.022	16	0	0	67.79	*Alphaproteobacteria*
54	0.022	16	0	0	51.84	None
55	0.02	15	3.6	0.7	53.99	None
56	0.014	10	0	0	61.93	None
57	0.016	9	0	0	57.86	*Alphaproteobacteria*
58	0.005	4	0	0	69.05	None

59	0.001	1	0	0	55.97	None
60	0.001	1	0	0	36.02	*Lutibacter*
61	0.001	1	0	0	52.06	*Vibrio*

Table S5. Individual ANG metagenome assembly manual bins

Assembly	Bin	Total size (Mbp)	Number of contigs	Percent completion	Percent redundancy	GC content (%)	Taxonomy
E7 individual	1	20	3,430	0	0	62.38	*Leisingera*
	2	1.96	716	60.4	7.9	62.88	*Opitutae*
	3	2.78	424	97.1	6.5	64.09	*Erythrobacter*
	4	10.5	9,478	0	0	61.52	*Alphaproteobacteria*
	5	9.37	917	0	0	58.29	*Alphaproteobacteria*
	6	4.82	97	99.3	1.4	56.34	*Mesorhizobium*
	7	8.9	1,436	77	64.7	58.07	*Alphaproteobacteria*
	8	4.99	919	96.4	27.3	40.78	*Muricauda*
	9	5.5	1,299	71.2	39.6	48.7	*Gammaproteobacteria*
	10	4	45	97.8	0.7	56.34	*Opitutae*
	11	10.6	1,656	3.6	1.2	53.87	*Alphaproteobacteria*
E8 individual	1	4.56	634	30.2	10.1	53.11	*Verrucomicrobia*
	2	2.2	346	2.5	0	50.65	*Gammaproteobacteria*
	3	2.2	413	64.7	26.6	53.96	*Gammaproteobacteria*
	4	2.26	678	46.8	0.7	50.22	*Gammaproteobacteria*
	5	4.91	151	100	5	44.77	*Vibrio*
	6	4.72	830	72.7	2.2	62.26	*Leisingera*
	7	3.32	14	56.8	0	62.81	*Ruegeria*
	8	5.82	173	94.2	39.6	60.24	*Ruegeria*
	9	2.8	572	31.7	0.7	57.75	*Alphaproteobacteria*
	10	5.25	218	76.3	28.1	57.78	*Alphaproteobacteria*
	11	6.4	812	32.4	12.2	54.51	*Alphaproteobacteria*
	12	11.2	1,215	0	0	57.13	*Rhizobium*
E16 individual	1	11.7	1,065	0	0	57.71	*Alphaproteobacteria*
	2	4.96	875	75.5	30.2	58.38	*Alphaproteobacteria*
	3	4.73	216	75.5	0.7	60.69	*Alphaproteobacteria*
	4	4.81	1,156	52.5	25.9	60.18	*Alphaproteobacteria*
	5	5.35	1,212	81.3	33.1	61.69	*Alphaproteobacteria*

	6	2.25	705	36	0.7	63.45	*Opitutae*
	7	7.91	1,202	87.1	40.3	63.02	*Leisingera*
	8	9.6	1,542	7.2	4.8	42.58	None
	9	12.3	1,183	0	0	53	None
	10	3.98	1,136	83.5	36.7	49.25	None
	11	10.4	1,655	93.5	69.1	54.29	*Alphaproteobacteria*
PD13 individual	1	5.1	4	96.4	1.4	52.5	*Verrucomicrobia*
	2	4.16	131	98.6	0	57.2	*Verrucomicrobia*
	3	0.54	161	2.2	0	50.9	None
	4	13	3,125	0	0	63.3	*Alphaproteobacteria*
	5	2.21	1,208	51.1	11.5	56.7	*Alphaproteobacteria*
	6	3.61	1,392	83.5	5	42.1	None

Table S6. AntiSMASH results from MAGs

MAG	Cluster type	Contig	Location start (bp)	Location end (bp)	Top homologous cluster hit	Percent of genes with similarity to top homologous hit	Most similar known cluster	Percent of genes with similarity to known cluster
Alphaproteo. CO Bin 41	bacteriocin-lanthipeptide	15	17,599	33,796	*Methylovulum psychrotolerans* HV10 M2	27	NA	NA
	NRPS-T1PKS	46	1	15,573	*Ramlibacter* sp. Leaf400	8	NA	NA
	betalactone	82	11,372	38,805	*Oceanicella actignis* DSM 24423	15	NA	NA
	HSl	82	167,411	188,295	*Cribrihabitans marinus* DSM29340	63	NA	NA
	Ectoine	85	1	6,645	*Ruegeria atlantica* CECT 4292	75	NA	NA
	terpene	86	16,950	37,726	*Ruegeria pomeroyi* DSS-3	85	NA	NA
	bacteriocin	86	469,384	478,867	*Limimaricola pyoseonensis* DSM 21424	13	NA	NA
	bacteriocin	86	560,173	570,160	*Phaeobacter* sp. CECT 7735	78	NA	NA
	bacteriocin	86	600,758	611,099	*Pelagibaca abyssi* JLT2014	4	NA	NA
	HSL	89	11,493	32,230	*Ruegeria profundi* ZGT108	83	NA	NA
	HSL	92	191,486	212,124	*Sulfitobacter pseudonitzschiae* DSM 26824	75	NA	NA
	bacteriocin	54	3,377	14,213	*Erythrobacter* sp. AP23	81	NA	NA

Erythro-bacter E7 Bin 3	terpene	67	1	4,308	*Erythrobacter* sp. AP23	23	Astaxanthin dideoxy-glycoside	75
Meso-rhizobium E7 Bin 6	terpene	44	1	15,861	*Mesorhizobium* sp. LC103	65	NA	NA
	NRPS	56	35,211	61,412	*Tistrella mobilis* KA081020-065	25	NA	NA
	HSL	66	30,136	50,774	*Mesorhizobium ciceri* biovar biserrulae WSM1271	68	NA	NA
	HSL	91	16,758	37,447	*Paramesorhizobium deserti* A-3-E	20	NA	NA
Opitutae E7 Bin 10	Other	4	16,761	57,453	*Verrucomicrobium* sp. BvORR106	7	NA	NA
	NRPS	5	53,045	98,906	*Bradyrhizobium* sp. BR10245	6	NA	NA
Verruco. PD13 Bin 1	NRPS-like	3	550,940	591,575	*Pseudomonas putida* F1	10	NA	NA
	NRPS	3	1,001,636	1,048,585	*Burkholderia ubonensis* MSMB1137	5	NA	NA
	NRPS	3	1,588,693	1,642,687	*Streptomyces* sp. ScaeMP-e83	3	NA	NA
	NRPS-like	3	2,667,171	2,709,030	*Synechococcus* sp. PCC 7336	4	NA	NA
Verruco. PD13 Bin 2	NRPS	46	23,461	51,798	*Lysobacter gummosus* 3.2.11	16	NA	NA
	Other	56	19,311	60,003	*Verrucomicrobium* sp. BvORR106	7	NA	NA
Verruco. CO Bin 35	NRPS-like	110	1	30,332	*Kutzneria albida* DSM 43870	2	NA	NA
	NRPS-like	174	1	2,661	*Synechococcus* sp. PCC 7336	2	NA	NA
	NRPS	179	1	30,429	*Streptomyces* sp. ScaeMP-e83	3	NA	NA

	NRPS	420	1	2,712	*Myxococcus virescens* DSM 2260	100	NA	NA
Vibrio E8Bin5	NRPS	37	1	12,065	*Vibrio tubiashii* ATCC19109	23	cupriachelin	11
	betalactone	49	77,355	99,876	*Vibrio litoralis* DSM 17657	24	NA	NA
	arylpolyene	99	7,903	51,489	*Vibrio europaeus* PP-638	92	APE Vf	90
	bacteriocin	104	14,690	25,547	*Vibrio europaeus* PP-638	100	NA	NA
Vibrio COBin40	bacteriocin	64	71,723	82,580	*Vibrio europaeus* PP-638	100	NA	NA
	NRPS	67	1	12,064	*Vibrio tubiashii* ATCC19109	23	Cupriachelin	11
	NRPS	112	1	3,765	*Nitrospira* sp. Nsp13	100	NA	NA
	NRPS	169	1	6,463	*Vibrio coralliilyticus* P1	9	NA	NA
	betalactone	176	192	22,232	*Vibrio litoralis* DSM 17657	24	NA	NA
	arylpolyene	184	7,904	51,490	*Vibrio europaeus* PP-638	92	APE Vf	85

Table S7. ANI analysis of *Verrucomicrobia* MAGs

	***Verrucomicrobia* PD13 Bin 1**	***Verrucomicrobia* PD13 Bin 2**	***Opitutae* E7 bin 10**	***Verrucomicrobia* CO bin 35**
***Verrucomicrobia* PD13 Bin1**	*	70.85	70.48	99.52
***Verrucomicrobia* PD13 Bin 2**	71.04	*	85.88	71.16
***Opitutae* E7 bin 10**	70.71	85.99	*	70.81
***Verrucomicrobia* CO bin 35**	99.9	71.2	70.83	*

ANI > 95% = Species level cutoff

Table S8. Predicted carbohydrate active enzymes in the *Verrucomicrobia* PD13 Bin 1 MAG

CAZy Family	Number of Genes	Predicted function	Number of signal peptides
AA2	1	Catalase-peroxidase	0
AA3	1	Glucose-methanol-choline (GMC) oxidoreductase	0
AA7	2	Glucooligosaccharide oxidase	0
CBM32	3	N-acetylgalactosamine 6-sulfate sulfatase	3
CBM38	3	Inulin binding/ iduronate-2-sulfatase	2
CBM4	1	Xylan/ β-1,3-glucan/ β-1,3-1,4-glucan/ β-1,6-glucan/amorphous cellulose binding	1
CBM47	3	Fucose binding / membrane-bound dehydrogenase	2
CBM48	1	1,4-alpha-glucan (glycogen) branching enzyme	0
CBM50	1	Peptidoglycan binding	1
CBM51	4	Galactose binding	4
CBM6	2	Cellulose/ β-1,4-xylan/ β-1,3-glucan/ β-1,3-1,4-glucan/ β-1,4-glucan binding	2
CBM67	6	Alpha-L-rhamnosidase	5
CBM9	1	Xylan binding	0
CE0	1	Carbohydrate sulfatase	1
CE1	6	Endo-1,4-beta-xylanase	2
CE10	14	Acetyl esterase / lipase	11
CE11	1	UDP-3-0-acyl N-acetylglucosamine deacetylase	0
CE12	1	L-glyceraldehyde 3-phosphate reductase	0
CE15	3	Acetylxylan esterase	1
CE16	1	Acetylxylan esterase	0
CE2	1	GDSL-like lipase/acylhydrolase domain protein	1

CE3	4	Arylesterase	2
CE4	2	Polysaccharide deacetylase family protein	1
CE6	2	Glycerophosphoryl diester phosphodiesterase	1
GH82	1	I-carrageenase	1
GH0	3	Myo-inositol 2-dehydrogenase	1
GH10	10	endo-1,4-β-xylanase	7
GH101	1	endo-α-N-acetylgalactosaminidase	0
GH106	2	α-L-rhamnosidase/ rhamnogalacturonan α-L-rhamnohydrolase	1
GH107	1	sulfated fucan endo-1,4-fucanase	0
GH109	37	α-N-acetylgalactosaminidase	13
GH110	4	α-galactosidase	3
GH115	1	xylan α-1,2-glucuronidase/ α-(4-O-methyl)-glucuronidase	0
GH116	2	β-glucosidase/ β-xylosidase/β-glucosylceramidase/ β-N-acetylglucosaminidase	0
GH117	18	α-1,3-L-neoagarooligosaccharide hydrolase/ α-1,3-L-neoagarobiase hydrolase	13
GH120	1	β-xylosidase	0
GH127	6	β-L-arabinofuranosidase/ 3-C-carboxy-5-deoxy-L-xylose hydrolase	3
GH128	1	β-1,3-glucanase	1
GH129	1	α-N-acetylgalactosaminidase	1
GH13	8	α-amylase/ pullulanase	2
GH130	4	1,4-beta-mannosyl-N-acetylglucosamine phosphorylase	0
GH136	3	lacto-N-biosidase	2
GH138	5	α-galacturonidase	3

GH14	1	β-amylase	0
GH140	3	Apiosidase	3
GH141	8	α-L-fucosidase/ xylanase	5
GH148	2	β-1,3-glucanase	2
GH149	1	β-1,3-glucan phosphorylase	0
GH150	2	l-carrageenase	1
GH151	4	α-L-fucosidase	2
GH16	14	Beta-glucanase precursor	9
GH17	2	glucan endo-1,3-β-glucosidase/ glucan 1,3-β-glucosidase/ licheninase/ ABA-specific β-glucosidase/ β-1,3-glucanosyltransglycosylase	0
GH2	23	β-galactosidase	13
GH20	1	β-1,6-N-acetylglucosaminidase/β-6-SO3-N-acetylglucosaminidase	0
GH26	2	Mannan endo-1,4-beta-mannosidase	1
GH27	2	α-galactosidase	0
GH28	6	polygalacturonase	1
GH29	26	α-L-fucosidase	21
GH3	2	β-glucosidase	0
GH30	3	endo-β-1,4-xylanase/ β-glucosidase/ β-glucuronidase/ β-xylosidase/ β-fucosidase/ glucosylceramidase/ β-1,6-glucanase/ glucuronoarabinoxylan endo-β-1,4-xylanase/ endo-β-1,6-galactanase/ [reducing end] β-xylosidase	2
GH31	4	α-glucosidase /α-xylosidase	1
GH32	2	Sucrose-6-phosphate hydrolase	0
GH33	12	Sialidase/ exo-alpha-sialidase	11
GH36	3	α-galactosidase	2

GH38	2	α-mannosidase	0
GH39	3	α-L-iduronidase/ β-xylosidase	1
GH4	2	α-glucosidase/ α-galactosidase	0
GH42	2	β-galactosidase	2
GH43	9	Xylan 1,4-beta-xylosidase	6
GH49	1	dextranase/ isopullulanase/ dextran 1,6-α-isomaltotriosidase/ sulfated arabinan endo-1,4-β-L-arabinanase	0
GH5	5	Endoglucanase C	5
GH50	10	β-agarase/ sulfatase	5
GH51	4	alpha-L-arabinofuranosidase	2
GH53	1	endo-β-1,4-galactanase/ sulfatase	1
GH55	1	exo-β-1,3-glucanase/ endo-β-1,3-glucanase	0
GH6	2	endoglucanase/ cellobiohydrolase	0
GH65	1	α,α-trehalase/ maltose phosphorylase/ trehalose phosphorylase/ kojibiose phosphorylase/ trehalose-6-phosphate phosphorylase/ nigerose phosphorylase/ 3-O-α-glucopyranosyl-L-rhamnose phosphorylase/ 2-O-α-glucopyranosylglycerol/ phosphate β-glucosyltransferase/ α-glucosyl-1,2-β-galactosyl-L-hydroxylysine α-glucosidase	0
GH73	1	Lysozyme/ mannosyl-glycoprotein endo-β-N-acetylglucosaminidase	0
GH77	2	4-α-glucanotransferase	0
GH78	4	α-L-rhamnosidase	2
GH81	1	endo-β-1,3-glucanase	1
GH82	2	I-carrageenase	1
GH88	1	d-4,5-unsaturated β-glucuronyl hydrolase	1

GH9	4	Endoglucanase/ endo-β-1,3(4)-glucanase / lichenase-laminarinase/ β-glucosidase/ lichenase / endo-β-1,3-1,4-glucanase/ exo-β-1,4-glucanase / cellodextrinase/ cellobiohydrolase/ xyloglucan-specific endo-β-1,4-glucanase / endo-xyloglucanase/ exo-β-glucosaminidase	2
GH92	1	Mannosyl-oligosaccharide α-1,2-mannosidase	1
GH94	2	Cellobiose phosphorylase	0
GH95	10	α-L-fucosidase/ α-1,2-L-fucosidase/ α-L-galactosidase	7
GH97	2	α-glucosidase	2
GH99	1	Glycoprotein endo-α-1,2-mannosidase/ mannan endo-1,2-α-mannanase	1
GT13	1	α-1,3-mannosyl-glycoprotein β-1,2-N-acetylglucosaminyltransferase	0
GT19	1	Lipid-A-disaccharide synthase	0
GT2	15	Bactoprenol glucosyl transferase	0
GT26	2	β-N-acetyl mannosaminuronyltransferase/ β-N-acetyl-mannosaminyltransferase/ β-1,4-glucosyltransferase/ β-1,4-galactosyltransferase	0
GT28	2	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	0
GT30	2	3-deoxy-D-manno-octulosonic acid transferase	0
GT35	1	Glycogen phosphorylase	0
GT4	20	Glycogen synthase/ ADP-glucose transglucosylase/ N-acetyl-alpha-D-glucosaminyl L-malate synthase	0
GT41	1	β-N-acetylglucosaminyltransferase/ N-β-glucosyltransferase	0
GT8	1	lipopolysaccharide α-1,3-galactosyltransferase/ lipopolysaccharide glucosyltransferase 1/ glycogenin glucosyltransferase/ inositol 1-	0

		α-galactosyltransferase/ homogalacturonan α-1,4-galacturonosyltransferase/ xylan α-glucuronyltransferase	
GT81	1	Glucosyl-3-phosphoglycerate synthase	0
GT94	1	GlcA-β-1,2-Man-α-1,3-Glc-β-1,4-Glc-α-1-PP-undecaprenol β-1,4-mannosyltransferase	0
PL0	1	Polysaccharide lyase	1
PL10	1	Pectate lyase	1
PL12	1	Heparin-sulfate lyase	1
PL13	1	Heparin lyase	1
PL14	1	Alginate lyase/ exo-oligoalginate lyase/ β-1,4-glucuronan lyase	0
PL15	1	Oligo-alginate lyase/ alginate lyase	0
PL25	1	Ulvan lyase	0
PL6	2	Alginate lyase	1
PL9	1	Pectate lyase/ Exo-polygalacturonate lyase/ thiopeptidoglycan lyase	0

Abbreviations: AA = auxiliary activity; CBM = carbohydrate binding module; CE = carbohydrate esterase; GH = glycosyl hydrolase; GT = glycosyl transferase; PL = polysaccharide lysase

*Predicted function is based on the RAST annotation or BLASTx result compared to the nt/nr database. Where an annotation could not be assigned, general predictions provided by the CAZy database are listed.

Chapter 4

Metatranscriptome analyses of accessory nidamental gland and egg jelly coat communities provide insights into symbiont primary and secondary metabolism

Introduction

The accessory nidamental gland (ANG) of *Euprymna scolopes* houses a bacterial consortium predominantly composed of *Alphaproteobacteria*, *Verrucomicrobia*, *Flavobacteriia*, and *Gammaproteobacteria* (69)(72). These bacteria are deposited into the jelly coat (JC) surrounding the yolk sac of bobtail squid eggs and are believed to provide defense against microbial infections (Appendix I). Organic extracts from ANG/JC isolates have demonstrated *in vitro* antimicrobial activity (85) (Chapter 1, Appendix I), suggesting that the community has the potential to produce small molecules *in vivo* that could play a role in egg defense. Previous analyses of ANG metagenomes predicted 255 secondary metabolite biosynthetic gene clusters (Chapter 3), some of which could potentially produce or regulate bioactive defensive compounds. To better understand which of these predicted gene clusters are expressed and the types of bacterial metabolism that occur in host conditions, metatranscriptomes were sequenced of the ANG, freshly laid (day 0) egg JCs, JCs from eggs mid-way through development (day 10 control), and JCs from eggs that were challenged with the fungal pathogen, *Fusarium keratoplasticum*, for half of embryogenesis (day 10 challenge) to determine if challenge with a pathogen increases expression of secondary metabolite genes.

Metatranscriptomics have been used to analyze community-wide bacterial gene expression in the presence of a host and associate functions in symbioses. In the wood-feeding higher termite, *Nasutitermes takasagoensis*, digestion of xylan, the main component of hemicellulose in the termite's diet, was found to be associated with *Spirochaetes* in the hindgut that express several xylanases (105). A time-course study sequenced metatranscriptomes from the human oral microbiome over the course of 24 hours to examine gene expression responses to pH changes in the environment (216). Many significant transcriptional changes were seen

between six and nine hours, as pH levels dropped from 5.5 to 4.3, including genes involved in pyruvate and lactate metabolism, two-component signaling pathways, iron acquisition, and biofilm formation. Of the 15 predicted secondary metabolite biosynthetic gene clusters with known products in this community, nine clusters for production of bacteriocins and non-ribosomal peptide synthetases (NRPSs) were differentially expressed across the different pH stages (216). These small molecules are thought to play a role in competition between strains during biofilm formation but may also have unexplored impacts on human health.

Dual RNA-sequencing can be used to examine gene expression of both a host and its symbionts at the same time. This technique has most often been used to study pathogenic symbioses, and can reveal which virulence factors are important in establishing an infection, while also revealing what host immune factors are expressed in response to the infection (217). Dual RNA-sequencing has less frequently been used to study beneficial symbioses, and can reveal what bacterial and host factors are important in establishing and maintaining colonization. Dual sequencing of the nematode, *Brugia malayi*, and its *Wolbachia* symbionts found many genes for energy production and transport in *Wolbachia* were upregulated during female nematode development, demonstrating that pathways such as purine biosynthesis are beneficial to the host (218). In this chapter, eukaryotic and prokaryotic transcripts from six ANGs were sequenced to examine what nutrient sources the bacteria are utilizing and which of these nutrients may be produced by the host.

Of the secondary metabolite gene clusters predicted in the ANG metagenome, 51.8% of the clusters were expressed in the ANG, 50.2% expressed in day 0 JCs, 51.8% expressed in day 10 control JCs, and 60% expressed in day 10 challenge JCs. The expressed clusters represented 15 metabolite classes, the majority of which were homoserine lactones, bacteriocins, and

terpenes. Non-ribosomal peptide synthetase core biosynthesis genes were the highest expressed in all samples, and gene expression was not significantly different between sample types. Examination of primary metabolism in the ANG/JC communities focused on expression of genes from the *Verrucomicrobia* PD13 Bin 1 MAG. This strain is expressing cell division genes at equal levels between all ANG/JC samples, and expressed genes to utilize sulfate, sialic acid, and fucose. Genes to synthesize fucosylated glycoproteins were expressed in ANG tissue, and stains for sulfated and sialylated mucins detected these glycoproteins in both the ANG tubules and JC of eggs.

Material and Methods

ANG and egg jelly coat (JC) sample collection

Adult Hawaiian bobtail squid, *Euprymna scolopes*, were collected from sand flats in Maunalua Bay, Oahu, Hawaii (21826'3.36"N, 157847'20.78"W) in July 2016, January 2017, July 2018, and September 2018. All bobtail squid were shipped to the University of Connecticut and maintained in aquaria with artificial seawater (Instant Ocean, Blacksburg, VA) and a 12 hr light / 12 hr dark cycle for at least one month. Six sexually mature females (mantle lengths = 23 – 27 mm, **Table S1**) were sacrificed following protocols approved by the UConn Institutional Animal Care and Use Committee (IACUC). The ANGs were dissected and divided in half using sterile forceps and razor blades treated with RNase Away (Ambion, Austin, TX) for one hour prior to use. Each ANG section was surface sterilized with 95% ethanol and rinsed in filter-sterilized squid Ringer's (FSSR, 530 mM NaCl, 25 mM $MgCl_2$, 10 mM $CaCl_2$, 20 mM HEPES, pH 7.5). One half of each ANG sample was placed in RNAlater (Invitrogen, Waltham, MA) and

stored at -80°C for future RNA extractions. The other half of each ANG was flash-frozen in liquid nitrogen and stored at -80°C for DNA extractions (see chapter 3).

While bobtail squid were maintained at the University of Connecticut, males and females were placed in the same tank to mate once every two weeks. After mating, females were placed in separate tanks to track the egg clutches laid by each female. Once an egg clutch was laid, it was removed from the female's tank and placed in a separate tank to keep track of the age of individual clutches. Clutches were numbered (**Table S2**) in reference to the order they were laid by all females and do not indicate the number laid by any one female. Egg clutches are recorded as "day 0" if they were sampled less than 24 hours after being laid. At least 30 eggs were sampled from a day 0 clutch laid by a female from which an ANG was also sampled.

Sampled eggs were gently removed from the entire egg clutch, surface sterilized with 95% ethanol, and rinsed with FSSR prior to dissection. Forceps were soaked in RNase Away for 1 h prior to use. To collect the jelly coat (JC) layer containing bacteria, the outer capsule of each egg was first removed and discarded. Forceps were used to tear a small hole in the JC to remove the yolk sac containing the embryo. Each JC was immediately placed into a bead-beating tube kept on dry ice. Due to the small volume of JC in each egg (approximately 6 mg), the JCs froze immediately when placed in these pre-chilled tubes. Tubes containing the frozen JC were stored at -80°C until RNA extraction. Preservation in RNAlater was not effective because this solution interfered with downstream homogenization and RNA extraction.

Egg Fungal Challenge Experiment

To determine the effect of fungal challenge on jelly coat bacterial transcription, eggs from five females (from which the ANG was also sampled, **Table S2**) were exposed to the

fungal pathogen, *Fusarium keratoplasticum* FSSC-2g, during development. For each sample, at least 60 day 0 eggs from one clutch per female were collected and divided into two groups: at least 30 eggs were placed in a control group and at least 30 eggs were in the challenge group and exposed to *F. keratoplasticum*. All eggs were maintained in 100 ml of filter-sterilized artificial seawater (ASW) in 250 ml polypropylene beakers with air stones. Beakers were covered with plastic wrap to prevent evaporation and changes in salinity. Water changes were performed every 2-3 days. All beakers containing eggs were maintained in a biological safety cabinet to prevent contamination. All non-biological materials placed in the safety cabinet were UV sterilized for one hour prior to initiating the experiment.

Prior to placement in the beakers, all eggs were rinsed six times with filter-sterilized ASW to remove any transient surface microbes. *F. keratoplasitcum* bud cells at 10^4 bud cells/ml were added to the challenge eggs on the first day of the experiment and after every water change. Bud cells were collected and quantified following previous protocols (Appendix I). At day 10 of development, JCs from both challenge and control eggs were collected and flash-frozen following procedures described above.

Accessory Nidamental Gland (ANG) RNA Extraction

ANG samples frozen in RNAlater were thawed on ice and placed into a bead-beating tube containing 0.01 mm and 0.5 mm diameter zirconia/silica beads (BioSpec Products, Bartlesville, OK) and 1.0 ml of TRIzol Reagent (Invitrogen, Carlsbad, CA). Samples were bead-beat at 3,500 rpm for two 45 s cycles with a 30 s pause using a PowerLyzer 24 Homogenizer (Qiagen, Germantown, MD). Beads were pelleted by centrifuging for 5 min, 16,000 x rcf, at 4°C. Supernatant was transferred to a new tube and RNA isolated following the TRIZol

manufacturer's protocol. RNA samples were treated with TURBO DNase (Invitrogen, Carlsbad, CA) for 1 h at 37°C. To confirm DNA was removed, no bands were seen on a 1% agarose gel when PCR amplifying RNA using the standard GoTaq Green Master Mix (Promega, Madison, WI) protocol with 16S rRNA gene universal primers, 30 cycles, and 55°C annealing temperature. RNA concentrations were measured using a Qubit RNA HS assay kit (Invitrogen, Waltham, MA) and RNA quality was checked using RNA ScreenTape on an Agilent TapeStation 2200 (Agilent Technologies, Santa Clara, CA).

To separate bacterial RNA from the host RNA, all samples were processed through a eukaryotic mRNA selection with cleaning and concentrating of leftover RNA. The NEBNext Poly(A) mRNA Magnetic Isolation Module (New England BioLabs, Ipswich, MA) was used to select eukaryotic mRNA and separate it from eukaryotic rRNA and total prokaryotic RNA. 0.9 ug – 4.2 ug of RNA (**Table S1**) were used as input, following the manufacturer's protocol except where noted below. At steps 16 and 20 of this protocol, after washing magnetic beads bound with eukaryotic mRNA, all supernatant was pooled and placed on ice until the protocol was completed. The mRNA recovered at this stage was prepared for sequencing as described below and hereafter referred to as the eukaryotic sample from each female.

A total of 500 µl of pooled supernatant containing total prokaryotic RNA and eukaryotic rRNA was cleaned using the Zymo RNA Clean and Concentrator Kit (Zymo Research, Irvine, CA). This concentrated RNA was treated with a 50:50 mixture of RiboZero bacteria and RiboZero Gold epidemiology rRNA removal probes (Illumina, San Diego, CA), to remove both prokaryotic and eukaryotic rRNA, following the manufacturer's protocol with ethanol precipitation. The RNA recovered from this procedure was prepared for sequencing as described below and hereafter referred to as the prokaryotic sample from each female, although some host

rRNA did remain after RiboZero treatment. All RNA samples were stored at -80°C until prepared for sequencing.

Egg Jelly Coat (JC) RNA Extraction

JCs frozen in bead-beating tubes were lyophilized in a FreeZone Plus 4.5 L cascade benchtop freeze dryer (Labconco Corp., Kansas City, MO) for 24 h at -84°C. Lyophilized samples were bead-beat with 0.01 mm and 0.5 mm diameter zirconia/silica beads (BioSpec Products, Bartlesville, OK) in 1.0 ml of TRIzol Reagent at 3,500 rpm for two 45 s cycles with a 30 s pause using a PowerLyzer 24 Homogenizer (Qiagen, Germantown, MD). Beads were pelleted by centrifuging for 5 min, 16,000 x rcf, at 4°C. Supernatant was transferred to a new tube and RNA isolated following the TRIZol manufacturer's protocol. RNA samples were treated with TURBO DNase (Invitrogen, Carlsbad, CA) for 1 h at 37°C. To confirm DNA was removed, no bands were seen on a 1% agarose gel when PCR amplifying RNA using the standard GoTaq Green Master Mix (Promega, Madison, WI) protocol with 16S rRNA gene universal primers, 30 cycles, and 55°C annealing temperature. RNA concentrations were measured using a Qubit RNA HS assay kit (Invitrogen, Waltham, MA) and RNA quality was checked using RNA ScreenTape on an Agilent TapeStation 2200 (Agilent Technologies, Santa Clara, CA). Due to low sample concentrations (below the Qubit RNA HS detection limit for all but one sample, **Table S2**), JC samples were not treated with RiboZero rRNA removal probes. Since the embryo was removed from all eggs prior to extraction, the samples were not expected to contain eukaryotic RNA.

Metatranscriptome Sequencing

Metatranscriptomic libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA). Due to the low concentrations of recovered RNA (**Table S1-S2**), all samples were concentrated to 5 µl using a ThermoSci Savant DNA 120 SpeedVac (ThermoFisher Scientific, Waltham, MA) on low heat (22°C). Concentrated samples were combined with 13 µl of the "Fragment, Prime, Finish Mix" (at the "Make RFP" step) and proceeded with library prep according to the manufacturer's protocol. Libraries were sequenced using the NextSeq 500 mid-output reagent kit v2 (2x150 cycles) (Illumina, San Diego, CA) at the Center for Genome Innovation (CGI), University of Connecticut.

Metatranscriptome read processing

Raw reads were quality filtered using Trimmomatic v.0.36 (121), removing any reads less than 36 bp with the following arguments: "LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36". Quality of reads was checked using FastQC v.0.11.7 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The rRNA reads were removed using SortMeRNA v.2.1b (219) and matching against the following SILVA rRNA databases: 90% ID bacterial 16S, 98% ID bacterial 23S, 95% ID archaeal 16S, 98% ID archaeal 23S, 95% ID eukaryotic 18S, 98% ID eukaryotic 28S, and the Rfam 5s and 5.8s databases. Remaining host transcripts were removed from the prokaryotic RNA samples using the methods described in (http://www.metagenomics.wiki/tools/short-read/remove-host-sequences). Prokaryotic non-rRNA reads were aligned to the *Euprymna scolopes* transcriptome (102) using BowTie2 v2.3.1 . Unmapped paired reads were filtered from bam files using samtools v.1.7. All prokaryotic reads

that did not align to the host transcriptome are referred to as the "host-filtered" reads. A flow chart of read processing and analyses can be found in **Figure S1**.

Metatranscriptome de novo assembly and analyses

All non-rRNA read pairs were used to create a de-novo metatranscriptome assembly with Trinity v.2.6.6. (220) with *in silico* read normalization according to depth of sequencing coverage. Trinotate v.3.0.2 (221) was used to annotate the de-novo assembly with TransDecoder v.3.0.1 (for predicting coding regions in transcripts), hmmer v.3.1b2 (domain identification), and blast.2.7.1 (transcript alignment). The de novo assembled contigs were analyzed using blastx against the Swiss-Prot database and using hmmscan against the Pfam database. Host-filtered prokaryotic reads were aligned to the de novo metatranscriptome using Bowtie2 v.2.3.1. A read counts matrix was generated using the "align_and_estimate_abundance.pl" script in Trinity v.2.6.6 with the RSEM estimation method. This counts matrix was used for differential expression analysis in the NOISeq v.2.26.1 package and TPM normalization in edgeR v.3.24.3, with scripts from (222), in R v3.5.2. Differentially expressed genes were visualized using the R package Manhattanly v.0.2.0.

Expression of *Verrucomicrobia* PD13 Bin 1 genes and ANG metagenome secondary metabolite biosynthetic gene clusters

Host-filtered prokaryotic reads were aligned to the *Verrucomicrobia* PD13 Bin 1 MAG (chapter 3) using CLC Genomics Workbench v.11 (Qiagen, Germantown, MD) with default alignment parameters. Host-filtered prokaryotic reads were also aligned to the secondary metabolite biosynthetic gene clusters predicted in the ANG metagenome coassembly (chapter 3,

Appendix III) using the same method as above. Aligned read counts were normalized across all metatranscriptome samples using the transcripts per kilobase million (TPM) method, which normalizes for gene length and then sequencing depth. Differential expression analyses were performed using the NOISeq v.2.26.1 package, with scripts from (222), in R v3.5.2. Differentially expressed genes were visualized using the R package Manhattanly v.0.2.0. Overlap in highest expressed genes between samples was visualized using the Venn diagram tool (http://bioinformatics.psb.ugent.be/webtools/Venn/). Heat maps were generated in Prism8 (GraphPad Software, San Diego, CA). To compare TPM values of each gene between samples in the heat maps, analyses of repeated measures were performed using the mixed-effects model with Tukey multiple comparison testing and Geisser-Greenhouse correction.

Metatranscriptome read-based analyses

Quality-filtered prokaryotic and eukaryotic reads were aligned to the GenBank non-redundant (nr) protein database release 92 with DIAMOND v.0.9.19 (195) and output format "100" for producing .daa files. Output files were imported into MEGAN6 community edition (223) using the "prot_acc2tax-Nov2018X1.abin" mapping file to extract taxonomic identities and the "acc2eggnog-Oct2016X.abin", "acc2interpro-June2018X.abin", and "acc2seed-May2015XX.abin" mapping files to respectively extract COG, Interpro, and SEED functional identifications for each read.

To focus analysis on the most dominant bacterial taxa associated with the ANG/JC communities, reads that were assigned to the *Alphaproteobacteria, Verrucomicrobia*, *Bacteroidetes*, and *Gammaproteobacteria* were individually exported from MEGAN. The proportions of reads assigned to each functional category for these taxa were averaged per

sample type (ANG/Day 0 JC/Day 10 JC control/Day 10 JC challenge) and analyses of repeated measures were performed using the mixed-effects model with Tukey multiple comparison testing and Geisser-Greenhouse correction.

Mucin Staining

The ANGs from 3 adult females were dissected as described above and fixed with 4% paraformaldehyde (PFA, diluted in FSSR) for 24 hours, rocking at 4°C. Five day 0 eggs from two clutches from two separate females were collected for fixation. A small incision was made in the outer capsule and JC to remove the yolk sac prior to fixation in either 4% PFA or Carnoy's solution (60% anhydrous ethanol, 30% chloroform, 10% glacial acetic acid) for 24 hours, rocking at 4°C. Fixed tissues were dehydrated with a series of absolute anhydrous ethanol (2x changes, 15 min), 50:50 anhydrous ethanol:xylene (2x, 30 min), and xylene (3x, 20 min). Tissues were embedded in paraffin and serial sectioned at 8 μm thickness. Sections were stored at 4°C prior to staining.

Sections were deparaffinized by melting at 60°C for 25 minutes and transferring through two changes of xylene for 10 min each. Sections were rehydrated by transferring through a graded ethanol series of 100% (2x), 90%, 80%, 70%, and 50% ethanol for five minutes each and five minutes in deionized (DI) water. Sections were stained with high-iron diamine (HID) to visualize sulfomucins and alcian blue to visualize sialomucins, following previous protocols (224). Sections were incubated with the HID solution (0.24% dimethyl meta-phenylenediamine dihydrochloride, 0.04% dimethyl para-phenylenediamine di-hydrochloride, and 1.7% $FeCl_3$) for 16 h at room temperature. Sections were rinsed with running DI water for 5 minutes, stained with 1% alcian blue (prepared in 3% acetic acid) for 10 min, and then rinsed again with running DI

water for 5 minutes. Sections were dehydrated by passage through the reverse graded ethanol series described above, cleared with two changes of xylene, and mounted with Permount (Fisher Chemical, Hampton, NH). Stained sections were visualized with a Zeiss Axiovert 200M at the Advanced Light Microscopy Facility at the University of Connecticut.

Results

Metatranscriptome sequencing

Sequencing of the prokaryotic ANG libraries yielded an average of 34.9 M ± 13.4 M raw read pairs and the eukaryotic ANG samples yielded 33.5 M ± 11.6 M raw read pairs (**Table S3**). For the egg jelly coat (JC) samples, day 0 JCs yielded 31.9 M ± 15.3 M raw read pairs, day 10 control JCs yielded 29.9 M ± 15.3 M raw read pairs, and the day 10 fungal challenge JCs yielded 37.3 M ± 21.2 M raw read pairs (**Table S3**). Quality filtering removed an average of 2.0 M ± 1.0 M read pairs from all samples. The average percent of rRNA read pairs removed was 13.3% ± 8.2% from prokaryotic ANGs, 1.2% ± 0.2% from eukaryotic ANGs, 59.2% ± 13.9% from day 0 JCs, 64.2% ± 9.2% from day 10 control JCs, and 68.0% ± 10.7% from day 10 challenge JCs (**Table S3-S4**). The amounts of remaining rRNA correspond to the sample treatments, as the eukaryotic ANG samples received poly-A mRNA selection, while the prokaryotic ANG sample was treated with rRNA removal probes, and the JC samples did not receive any rRNA removal treatments. After host-filtering the non-eukaryotic, non-rRNA samples, there were 21.5 M ± 8.7 M prokaryotic ANG read pairs, 20.2 M ± 14.1 M day 0 JC read pairs, 20.4 M ± 7.6 M day 10 control JC read pairs, and 26.5 M ± 18.8 M day 10 challenge JC read pairs (**Table S5**).

Secondary metabolite genes are expressed in the ANG and egg jelly coats (JCs)

Predicted secondary metabolite biosynthetic gene clusters from the ANG metagenome co-assembly (chapter 3, Appendix III), contain both the predicted core biosynthetic genes and neighboring genes which may be involved in tailoring, transport, or transcriptional regulation of secondary metabolites. A secondary metabolite gene cluster was reported as expressed if all the biosynthetic genes in that cluster had expression values. Overall, 51.8% of the clusters were expressed in the ANG metatranscriptomes, 50.2% expressed in the Day 0 JCs, 51.8% expressed in Day 10 control JCs, and 60% expressed in Day 10 fungal challenged JCs (**Fig. 1, Appendix IV**). Within all the samples, the homoserine lactones (HSLs) had the most clusters expressed (24-26% of clusters), followed by bacteriocins (15%) in the ANG and terpenes (15-16%) in all JC samples (**Fig. 1, Fig. 3A, Fig. 2A, Fig. 5A**). Other classes of secondary metabolites that had clusters expressed in all metatranscriptome samples include an acyl amino acid (**Fig. 7A**), ectoines (**Fig. 2B**), T1 PKS-NRPS (**Fig. 7B**), T1-PKS (**Fig. 5C**), T3-PKS (**Fig. 6A**), thiopeptides (**Fig. 5B**), siderophores (**Fig. 6B**), aryl polyenes (**Fig. 4A**), NRPSs (**Fig. 3B**), and "other" compounds (as categorized by anitSMASH, **Fig. 4**). The one nucleoside cluster was only expressed in the ANG (**Fig. 7D**), while cyanobactin was expressed in the Day 10 JCs (**Fig. 7C**) and T3-PKS-arylpolyene was detected in only one day 10 JC fungal challenge metatranscriptome (**Fig. 6C**).

When examining the top ten highest expressed genes in these clusters, considering both biosynthetic and neighboring genes, six genes were highly expressed in all ANG and JC samples (**Fig. 8A, Table S6**). These genes were an RNA polymerase sigma factor 70, *rpoD* (NRPS cluster 77, gene 7), DUF3520 domain-containing protein (NRPS cluster 77, gene 8), undecaprenyl/decaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferases

(NRPS cluster 220, gene 22/ NRPS cluster 194, gene 15), PEP-CTERM sorting domain-containing protein (NRPS cluster 220, gene 30), and a hypothetical protein (NRPS cluster 229, gene 16) (**Tables S6**).

Examining only the core biosynthetic genes in the clusters, there were only 3 genes that were in the top ten highest expressed for all ANG and JC samples (**Fig. 8B, Table S7**). These genes were two undecaprenyl/decaprenyl-phosphate (UDP) alpha-N-acetylglucosaminyl 1-phosphate transferases (NRPS cluster 220, gene 22 / NRPS cluster 194, gene 15) and one AMP-binding protein (Other cluster 207, gene 1) (**Table S7**). These clusters have top homologous cluster hits to *Paenibacillus alvei* (cluster 220), *Nodularia spumigena* (cluster 194), and *Microcoleus chthonoplastes* (cluster 207), but are on contigs assigned to the *Alphaproteobacteria*, *Betaproteobacteria*, or no assignment, respectively. The (UDP) alpha-N-acetylglucosaminyl 1-phosphate transferases had TPM normalized expression values ranging from 25,337 (day 10 challenge JC) up to 138,747 (day 0 JC) (**Fig. 3B**). The AMP-binding protein had TPM values ranging from 1,653 (day 0 JC) up to 3,741 (ANG) (**Fig. 3B**). In comparison, the *rpoD* gene in NRPS cluster 77 had expression values ranging from 88,613 (day 10 challenge JC) up to 106,642 (day 0 JC). Other biosynthetic genes that were highly expressed include six autoinducer synthases with closest blastp hits to *Ruegeria* spp (HSL clusters), two polysaccharide biosynthesis/export proteins (NRPS clusters), a non-ribosomal peptide synthetase (NRPS cluster), a sugar transferase (T1PKS), a lipopeptide synthetase (other), an adenylation domain containing protein (NRPS), a FabA-like protein (arylpolyene), and a hypothetical protein (other)(**Tables S7-S8**).

Comparing the average TPM values for core biosynthetic genes between samples (**Fig. 2-7**), only 1 comparison in the HSLs and 4 comparisons in the NRPSs were statistically significant.

For the HSLs, the SDR family NAD(P)-dependent oxidoreductase was higher in day 10 control JCs compared to the ANGs ($p = 0.0207$) (**Table S9**). Despite higher concentrations of bacteria in the ANG compared to egg JCs, none of the autoinducer synthases predicted in the ANG metagenome were expressed significantly higher in the ANG. This could be due, however, to high levels of variability between replicates, or we did not capture expression of the exact HSL clusters predicted in the metagenome. In the NRPSs, one undecaprenyl/decaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase (heat map row 1) was expressed higher in day 0 JCs compared to day 10 control JCs ($p = 0.0198$) and day 10 challenge JCs ($p = 0.0109$) (**Table S9**). A second undecaprenyl/decaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase (heat map row 2) was also expressed higher in day 0 JCs than in day 10 control JCs ($p = 0.0495$) and day 10 challenge JCs ($p = 0.0295$) (**Table S9**). Although some TPM values were considered significantly different for individual gene comparisons using the mixed-effects model and Tukey multiple comparison testing, overall, there were no significantly differentially expressed genes ($p > 0.05$) between any of the samples when analyzed with NOISeq (**Fig. S2**).

Verrucomicrobia PD13 Bin 1 gene expression

Expression of the *Verrucomicrobia* PD13 Bin 1 MAG was determined by mapping all metatranscriptome reads to the MAG and calculating normalized TPM expression values. Of the 4,095 coding sequences in this MAG, 4,003 had at least one TPM value. Of the top 20 highest expressed genes, 11 genes were shared between all ANG and JC samples (**Fig. 9, Table S10**). These genes were an MFS transporter, Yqge/AlgH family protein, glycosyl hydrolase, PBS lyase, mandelate racemase/muconate, sugar phosphate isomerase/epimerase, thiol peroxidase

(Bcp-type), carbohydrate-binding protein, and two proteins with unknown function. Genes that were most highly expressed in just JC samples were a response regulator, RNA polymerase sigma factor RpoE, GNAT family N-acetyltransferase, and 7 hypothetical proteins. Genes that were most highly expressed in the ANG include an acyl carrier protein, peptide transporter, and a hypothetical protein.

Carbohydrate active genes

Genes assigned as carbohydrate active enzymes made up one of the categories with the highest expressed genes. The highest expression was of a hypothetical protein assigned to the glycosyl hydrolase (GH) family 81, with an average TPM value of 24,002.4 in day 10 fungal challenge JCs (**Fig. 10F, Appendix V**). The GH family 81 enzymes are endo-β-1,3-glucanases. The next highest expressed gene was a hypothetical GH117 family protein, with an average TPM value of 21,467.5 in day 0 JCs (**Appendix V**). GH family 117 proteins are α-1,3-L-neoagarooligosaccharide hydrolases or α-1,3-L-neoagarobiase hydrolases. The third highest expressed gene was a hypothetical carbohydrate binding domain (CBM) 51 family protein, which bind galactose, with a TPM value of 6,085.9 in day 10 challenged JCs (**Fig. 12A**). All genes assigned to the 3 auxiliary activity (AA) families, 11 carbohydrate esterase (CE) families, and 9 polysaccharide lyase (PL) families were expressed in all ANG and JC samples (**Fig. 12A**). Of the 25 carbohydrate-binding domain family genes, 24 were expressed in the ANG (CBM4 not detected) and 23 were expressed in all of the JC samples (CBM4 and one of 5 CBM67 genes not detected). Of the 306 GH family genes, 305 were expressed in the ANG and day 0 JCs (GH109 not detected), 303 were expressed in day 10 control JCs (GH109, GH150, GH99 not detected), and 304 were expressed in day 10 challenge JCs (GH109 and GH150 not detected). Of the GT

family proteins, all 48 were expressed in the ANG, day 0 JCs, and day 10 challenge JCs, and 47 were expressed in day 10 control JCs (one of 15 GT2 genes not detected).

Cell division and RNA polymerase sigma factors

All cell division genes detected in the *Verrucomicrobia* PD13 Bin 1 MAG were expressed in all ANG and JC metatranscriptomes, except for one extra copy of a *ftsH* gene (**Fig. 10D**). The other *ftsH* copy was the highest expressed gene in this category, with TPM values ranging from 427.8 ± 204.1 in day 10 control JCs up to 636.6 ± 204.2 in ANGs (**Appendix V**). The second highest expressed gene in this category was *ftsI* (176.4 ± 267.9 TPM in day 10 control JCs up to 313.1 TPM in ANGs). No genes in this category had significantly different TPM values.

The sigma factors expressed in this MAG were four copies of *rpoE*, three copies of *rpoD*, three copies of *rpoN*, and one copy of *fliA*, the flagellar operon sigma factor (**Fig. 10A**). The highest expressed sigma factor was *rpoE* (row 6, **Fig. 10A, Appendix V**), ranging from 5,276.2 ± 4,967.5 TPM in ANGs up to 11,492.5 ± 5,758.3 TPM in day 10 challenge JCs. None of the sigma factors had significantly different TPM values between the ANG and JC samples.

Motility and chemotaxis

All flagellar biosynthesis and chemotaxis gene present in the MAG were expressed in all ANG and JC samples at relatively low levels, ranging from 3.6 TPM (*flgH* in day 10 challenge JCs) up to 59.6 TPM (*fliF* in day 0 JCs) (**Fig. 10E, Appendix V**). The highest expressed gene in this category was the flagellar M-ring gene, *fliF*, in ANG and day 0 JCs, while the methyl-accepting chemotaxis protein was highest in the day 10 JCs. The flagellar basal-body rod gene,

flgG, had a significantly higher average TPM value in ANGs compared to day 10 control JCs (p = 0.0157). The flagellum biosynthesis repressor protein, *flbT*, was significantly higher in ANGs versus day 10 challenge JCs (p = 0.0084) (**Table S11**).

Type IV pili

The *Verrucomicrobia* PD13 Bin 1 MAG had one copy of the type IV pili assembly gene, *pilC*, and four copies of the type IV pili ATPase, *pilB*, both of which were expressed in all samples (**Fig. 10C**). The *pilC* gene had the highest expression of these six genes, ranging from 159.9 ± 75.9 TPM in day 10 control JCs up to 967.8 ± 1425.2 TPM in ANGs (**Appendix V**). This gene was significantly higher in day 0 JCs compared to day 10 control JCs (p = 0.0417) and day 10 challenge JCs (p = 0.0161). Two of the *pilB* copies were also significantly higher in day 0 JCs compared to day 10 challenge JCs (p = 0.0241, 0.0382) (**Table S11**).

Transport genes

Transporters for ammonium, acetate, D-glycerate, D-xylose, ferrous iron, L-rhamnose, N-acetylglucosamine, oligopeptides (*OppBCDF*), sialic acid, galactose, glucose, choline, sulfate, and xyloside were expressed in all samples (**Fig. 12C**). Other transport systems expressed in all samples included the TonB-dependent receptors, TRAP-type transporters, twin-arginine translocation proteins (*tatAC*), and 49 of the 58 predicted ABC transporters. A biopolymer transport protein (ExbD/TolR) had the highest expression in ANGs (1,691.2 ± 994.9 TPM) and day 0 JCs (1,527.9 ± 1,056.6 TPM), while an ABC-type antimicrobial peptide transport system was highest in day 10 control JCs (978.2 ± 443.5 TPM) and day 10 challenge JCs (961.6 ± 573.0 TPM) (**Appendix V**).

Four transporters had significantly higher average TPM values in ANGs than day 0 JCs, including an ABC-type nitrate/sulfonate/bicarbonate transporter ATPase (p = 0.0273), an RND efflux inner membrane transporter (p = 0.016), and two uncharacterized permeases (p = 0.04332, 0.0459) (**Table S11**). Five transporters were significantly higher in day 0 JCs than in day 10 control JCs, including a L-rhamnose-proton symporter (p = 0.0043) and four uncharacterized permease/transport proteins. The oligopeptide transport gene, *oppD*, and a sodium/sulfate symporter were expressed higher in day 10 control JCs compared to day 10 challenge JCs (p = 0.0106, p = 0.0442, respectively) (**Table S11**).

Sulfur utilization and sulfatases

All genes detected for assimilative sulfate reduction were expressed in all ANG and JC samples (**Fig. 11C**). These genes included sulfate permeases (50.5 TPM in day 10 challenge JCs up to 97.0 in ANGs), sulfate adenylyltransferases (37.5 TPM in day 0 JCs up to 142.6 TPM in ANGs), adenylylsulfate kinase (38.2 TPM in day 10 control JCs up to 46.6 TPM in ANGs), phosphoadenylyl-sulfate reductase (40.7 TPM in ANGs up to 416.8 TPM in day 10 control JCs), and sulfite reductases (46.3 TPM in day 0 JCs up to 144.1 TPM in day 10 control JCs). The phosphoadenylyl-sulfate reductase had higher TPM values in ANGs compared to day 0 JCs (p = 0.0456) and day 10 control JCs (p = 0.0291). The cysteine synthase was also higher in ANGs compared to day 0 JCs (p = 0.0289) (**Table S11**). Of all the sulfur utilization genes, a cysteine synthase (row 5, **Fig. 11C**) had the highest expression, ranging from 648.4 ± 699.2 TPM in ANGs up to 4,711.2 ± 4,560.9 TPM in day 10 control JCs.

All 255 previously described sulfatases in the *Verrucomicrobia* PD13 Bin 1 MAG (chapter 3) were expressed in all ANG and JC samples, including some additional putative

sulfatases (**Fig. 11A-B**). Expression of sulfatases in the ANG ranged from 3.0 ± 4.1 TPM to 497.8 ± 374.0 TPM, in day 0 JCs from 1.4 ± 0.9 TPM to 909.8 ± 609.0 TPM, in day 10 control JCs from 2.0 ± 2.8 TPM to 761.4 ± 352.5 TPM, and in day 10 challenge JCs from 3.2 ± 3.7 TPM to 848.1 ± 282.8 TPM (**Appendix V**). One arylsulfatase was expressed higher in ANGs than day 0 JCs, 8 arylsulfatase/4 sulfatases were higher in day 0 JCs than day 10 challenge JCs, 5 arylsulfatases/3 sulfatases were higher in day 0 JCs than day 10 control JCs, and 2 arylsulfatases were higher in day 10 control JCs than day 10 challenge JCs (**Table S11**).

Fucose utilization

All 24 copies of the alpha-L-fucosidase gene were expressed in all ANG and JC samples, ranging from 8.5 ± 5.4 TPM (day 10 control JCs) up to 736.2 ± 603.0 TPM (day 10 challenge JCs) (**Fig. 12B**). Overall, fucose utilization genes were relatively low compared to other carbohydrate active enzymes. Fucose permease expression ranged from 23.7 ± 10.8 TPM (ANGs) up to 133.7 ± 62.7 TPM (ANGs), L-fucose isomerase expression ranged from 92.0 ± 56.2 (day 10 control JCs) up to 408.4 ± 211.8 TPM (ANGs), L-fucose kinase expression ranged from 33.4 ± 11.7 (day 10 challenge JC) up to 38.0 ± 12.5 (ANG), and L-fucose mutarotase ranged from 37.7 ± 16.2 (day 0 JC) up to 44.6 ± 17.1 TPM (day 10 challenge JC) (**Appendix V**). One alpha-L-fucosidase (row 24, **Fig. 12B**) was expressed higher in ANGs compared to day 0 JCs ($p = 0.042$) and day 10 control JCs ($p = 0.0166$) (**Table S11**). A different alpha-L-fucosidase (row 15, **Fig. 12B**), however, was expressed higher in day 10 control JCs compared to ANGs ($p = 0.021$). The L-fucose isomerase (row 27, **Fig. 12B**) was expressed higher in ANGs compared to day 0 JCs ($p = 0.0486$), day 10 control JCs ($p = 0.032$), and day 10 challenge JCs ($p = 0.0432$). Overall, L-fucose isomerase (row 27) had the highest expression in ANGs (408.4 ± 211.8 TPM),

while alpha-L-fucosidase (row 15) had the highest expression in all JC samples (286.5 ± 214.8 TPM in day 0 JC, 569.1 ± 388.1 TPM in day 10 control JCs, 736.2 ± 603.0 TPM in day 10 challenge JCs).

Sialic acid utilization

Similar to fucose utilization, sialic acid utilization genes were expressed in all ANG and JC samples, although at lower levels than most carbohydrate active enzymes (**Fig. 10B**). Three sialic acid-specific 9-O-acetylesterases were expressed, with one (row 5, **Fig. 10B**) being highest in all samples (52.6 ± 20.5 TPM in ANGs, 51.8 ± 30.7 in day 0 JCs, 38.0 ± 23.8 in day 10 control JCs, 43.7 ± 25.7 in day 10 challenge JCs) (**Appendix V**). A sialidase was expressed, ranging from 22.6 ± 10.2 TPM (day 10 control JCs) up to 37.0 ± 13.5 TPM (ANGs), as well as two copies of sialic acid transporters, ranging from 12.6 ± 6.4 TPM (day 0 JCs) up to 25.4 ± 13.4 TPM (ANGs). One sialic acid transporter (row 2) was expressed higher in day 10 challenge JCs than in day 0 JCs ($p = 0.0354$) (**Table S11**).

Differentially expressed genes in the *Verrucomicrobia* PD13 Bin 1 MAG

Differential expression analysis of the reads aligned to the *Verrucomicrobia* PD13 Bin 1 MAG showed nine differentially expressed genes amongst the three main sample comparisons of interest (**Fig. 13, Table S12**) and an additional 23 genes differentially expressed in other ANG and JC comparisons (**Fig. S3, Table S12**). In the comparison of ANGs versus day 0 JCs, a cell wall hydrolase and a fasciclin domain-containing protein were higher in day 0 JCs (**Fig. 13A**). In the comparison of day 0 JCs versus day 10 control JCs, 4 genes were higher in day 0 JCs: undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase, GHKL

domain-containing protein, and two hypothetical proteins (**Fig. 13B**). In the comparison of day 10 control JCs versus day 10 challenge JCs, 2 hypothetical proteins were higher in control eggs, while one hypothetical protein was higher in challenge eggs (**Fig. 13C**). All three of these hypothetical proteins did not have significant matches in the non-redundant NCBI database. The inverted nature of this plot was likely driven by either low expression values or variability between individual females.

Sulfomucins and sialomucins present in ANGs and egg JCs

After expression of genes associated with mucin utilization were observed in the metatranscriptomes, staining was performed on ANG and JC paraffin-embedded sections to determine which mucins were present in these samples. In the same ANG section, opposite mucin staining patterns were observed in different clusters of tubules, where the lumen of some tubules contained sialomucins while the tubule epithelia contained sulfomucins (**Fig. 14A**) while in other areas the opposite staining pattern was observed (**Fig. 14B**). In egg JC sections, Sulfomucins and sialomucins were observed in distinct layers, with sialomucins towards the outer capsule (**Fig. 14C-D**).

Adult *E. scolopes* ANG tissue expresses genes for synthesis of fucose

The *Euprymna scolopes* transcriptome (102) was searched for the presence of fucose synthesis genes. Isoforms for a GDP-L-fucose synthase (synthesizes fucose), a GDP-fucose transporter (transfers fucose to the golgi complex), and an alpha-1,6-fucosyltransferase (attaches fucose to glycopeptides) were seen in ANG tissues (**Fig. 15**). These genes were expressed 1.4 up

to 4 $\log_2$ fold change higher in the ANG than in other adult tissues, including the brain, gills, hemocytes, light organ, skin, and eyes.

Discussion

Secondary metabolite biosynthesis genes are expressed in the ANG and JCs

Due to the difficulty of extracting high-quality RNA from JC samples, and the separation of prokaryotic and eukaryotic RNA in ANG samples, low amounts of RNA were used as input in sequencing. Despite these low concentrations, an average of 33.5 M ± 13.4 M raw read pairs were obtained for all sample types. The separation of prokaryotic/eukaryotic RNA and rRNA removal performed on ANG samples did result in lower rRNA concentrations remaining after sequencing (2.7% up to 22.1% in prokaryotic ANGs, 0.98% up to 1.5% in eukaryotic ANGs) compared to the egg JC samples (36.6% up to 81.1%). After host-filtering of prokaryotic samples, an average of 21.9 M ± 12.3 M read pairs from all samples were used for mapping to reference gene clusters.

Of the predicted secondary metabolite gene clusters, 50-60% of core biosynthetic genes were expressed in all ANG and JC samples. Although only core genes were the focus of this analysis, many of the accessory genes needed to modify the core structures to complete synthesis of the small molecules were also expressed. The classes of metabolites produced were similar throughout all samples, with several non-ribosomal peptide synthetase clusters showing the highest expression. These clusters did not have predictions for the base structures that would be produced or the complete small molecule that may be produced by these pathways. The clusters also had low similarity to other known clusters, suggesting that these genes in the ANG may represent novel biosynthetic pathways or the metagenomic assemblies require deeper sequencing

to better assemble these gene pathways (disambiguate repetitive regions in these biosynthetic genes). Although expression of these clusters was not significantly different between samples, gene 22 of cluster 220 and gene 15 of cluster 194 both had the highest average expression in day 0 JC samples. The high degree of expression variability between females is likely the reason that no genes were significantly differentially expressed between any of the samples using NOISeq analysis.

Both of the highest expressed core biosynthetic secondary metabolite genes encode undecaprenyl-phosphate (UDP) alpha-N-acetylglucosaminyl (GlcNAc) 1-phosphate transferases. This protein catalyzes the transfer of the GlcNAc-1 phosphate moiety from UDP-GlcNAc onto a carrier lipid, undecaprenyl phosphate (C55-P). This transfer yields GlcNAc-pyrophosphoryl-undecaprenyl (GlcNAc-PP-C55), the first lipid-linked intermediate involved in production of enterobacterial common antigen, and is added to sugars to form the O-antigen lipopolysaccharide (LPS)(225).

Amongst all the samples, the homoserine lactones (HSLs) were the group with the most core biosynthetic genes expressed and six autoinducers were among the highest expressed genes. These autoinducers were primarily from *Rhodobacterales* strains, the most abundant group of bacteria in the ANG/JC communities. It was surprising that these genes were not differentially expressed between ANG and JC samples, due to the decrease in population density upon deposition into the eggs (72). Again, due to the variability that was seen between females, further research will need to clarify if this trend holds true in a larger sample size. Future research is needed to determine if homoserine lactones play a role in regulating gene expression of other secondary metabolite biosynthetic pathways.

Expression of Verrucomicrobia PD13 Bin 1 MAG metabolism genes

The *Verrucomicrobia* PD13 Bin 1 MAG had the most contiguous assembly of all the *Verrucomicrobia* MAGs assembled from the ANG metagenomes (Chapter 3) and was thus chosen as a reference for metatranscriptome mapping. Among the most highly expressed genes in all samples were two carbohydrate active enzymes, a major facilitator superfamily (MFS) transporter, mandelate racemase/muconate, a sugar phosphate isomerase/epimerase, and a thiol peroxidase (Bcp-type).

The majority of the cell division genes detected in the MAG were expressed at similar low levels between the ANG and JC samples. Bacteria in the JCs have been observed via TEMs to undergo 3-4 cell divisions over 21 days of egg development (72). While bacteria in the ANG are densely packed, their division rates are unknown. Rapid growth may only occur in the ANG during colonization or after the contents of a tubule have been emptied to colonize a JC during egg laying. At other times, the host may be regulating symbiont division to prevent an overgrowth of the community. In the light organ symbiosis, *E. scolopes* expels 95% of the *V. fischeri* cells each morning (226), but it is unknown if any venting occurs in the ANG symbiosis as well. Other hosts have been shown to regulate symbiont growth through use of antimicrobial peptides, such as those produced in weevils that inhibit cell division of endosymbionts in their bacteriocytes (227). Thus, it is plausible that the low expression levels of cell division genes reflect low division rates in both the ANG and egg environments.

Flagellar biosynthesis and chemotaxis genes were also expressed at low levels in both the ANG and JCs, with the basal body rod gene expressed significantly higher in the ANG compared to day 10 control JCs. While previous TEMs of *E. scolopes* ANGs were not stained to detect flagella (69), there is a report in *D. pealei* ANGs that approximately 10% of symbionts were

motile when homogenates were viewed using phase-contrast microscopy (67). The observed motility was not described with any particular symbiont morphology or taxonomy. Although the *Verrucomicrobia* have not yet been cultured, future chemotaxis assays or characterization of the ANG microenvironment may reveal which chemoattracts the *Verrucomicrobia* can respond to. Although motility by flagella has been described in other *Verrucomicrobia* (228)(229), there are no descriptions of known *Verrucomicrobia* chemoattractants.

Mucin staining revealed that sulfomucins and sialomucins are present in the ANG and eggs, and host expression data showed that fucosylated mucins may also be present in the ANG. Genes to cleave fucose, sialic acid, and sulfatase were expressed in *Verrucomicrobia* PD13 Bin 1, as well as transporters to bring these nutrients into the cell and continue their metabolism, leading to the predicted model in **Figure 16** that this symbiont is utilizing host-derived glycoproteins. Mucin staining also revealed a segregation of sulfomucin-dominant and sialomucin-dominant tubules in the same ANG sections. Fluorescence in situ hybridization (FISH) of ANG sections has previously shown that different bacterial taxa are segregated into different ANG tubules (69). The differential mucin staining suggests that the host may be provisioning different nutrient sources in the tubules that could create niche environments during colonization. A previous study in humans showed that differences in bacterial community composition along different sites of the colon were correlated with the total abundance of sulfomucin and sialomucin goblet cells (230). Higher amounts of these acidic mucins can lead to higher viscosity and acidity that may select for different microbes. In the light organ symbiosis, the host releases mucin on the ciliated appendages 1-2 hours after hatching, which provides a substrate that *V. fischeri* attaches to and forms aggregates before entering the pores for colonization (224). Mucin may also play a role in the attachment of symbionts during

colonization of the ANG tubules. Further characterization of the ANG microenvironments will reveal if other nutrient sources are also segregated between different tubules.

Transporters for glucose, galactose, L-rhamnose, D-glycerate, D-xylose, xyloside, and N-acetylglucosamine were all expressed at similar levels, suggesting that *Verrucomicrobia* has access to and can utilize a range of sugars in both the ANG and eggs. Some of the highest expressed glycoside hydrolases included the alpha-1,3-L-neoagarooligosaccharide hydrolase [GH117 family, agarose breakdown, (231)], endo-beta-1,3-glucanase (GH81/GH16, breaks down beta-glucans, a common energy reserve of algae), beta-glucosidase (GH116, acts on beta-D-glucosides to release glucose), alpha-N-acetylgalactosaminidase [GH109, cleave glycans from mucin backbones (199)], alpha-L-arabinofuranosidase (GH51, breakdown of arabinoxylan, component of plant hemicellulose), alpha-amylase (GH13, cleaves alpha-bonds in polysaccharides), xylan alpha-1,2-glucuronidase (GH115, acts on xylan polysaccharides), and alpha-L-fucosidase [GH95/GH29, cleave fucose from O-glycosidic chains linked to galactose or N-acetyl-glucosamine in mucins (200)]. Overall, these enzymes are associated with the utilization of plant-derived polysaccharides or mucins. While the origin of mucins can be directly associated with the bobtail squid host, plant-derived sugars may be associated with the host's diet. Although bobtail squid are predators of live shrimp, the shrimp may be feeding on algae in seawater. As well, the ANG bacteria may be exposed to algal sugars through the bobtail squid's respiration. The mantle cavity, which contains the majority of the bobtail squid's organs (including the ANG), is exposed to the surrounding seawater as water is pumped in from the mantle's openings, over the gills, and out through the siphon. The JC bacteria may be exposed to algae in surrounding seawater during the course of embryo development. There are no reports

regarding settlement of algae on squid eggs, but exposure to algae in this environment is highly likely.

This analysis has focused on just one of the MAGs recovered from the ANG metagenome. Mapping transcripts to MAGs from other taxa, or examining the results of read-based analyses for different taxa, will reveal if utilization of mucin and algal polysaccharides is unique to the *Verrucomicrobia* symbionts. If sequestration of different nutrient sources does occur in the ANG tubules, niche specialization can be hypothesized to occur amongst the different symbionts. Adaptations to utilize different nutrients amongst the different members of the ANG community would explain the taxa partitioning observed in the ANG community.

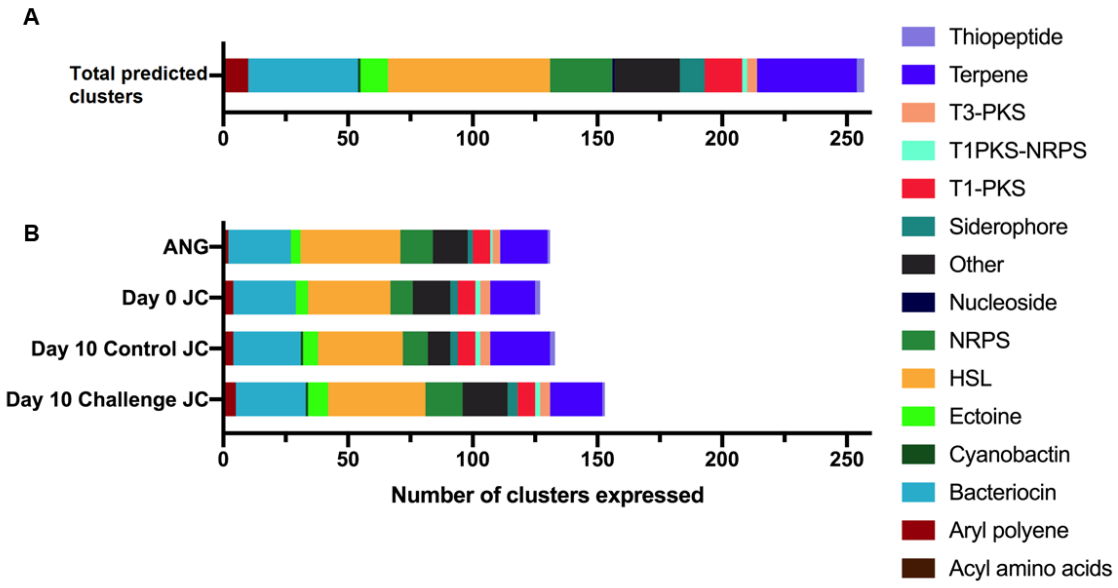


Figure 1. Number of expressed secondary metabolite biosynthetic gene clusters in ANG and JC metatranscriptomes. (A) Total number of predicted secondary metabolite gene clusters in the ANG metagenome co-assembly (see chapter 3). **(B)** Number of gene clusters expressed in the ANG, day 0 JCs, day 10 control JCs, and day 10 challenge JCs.

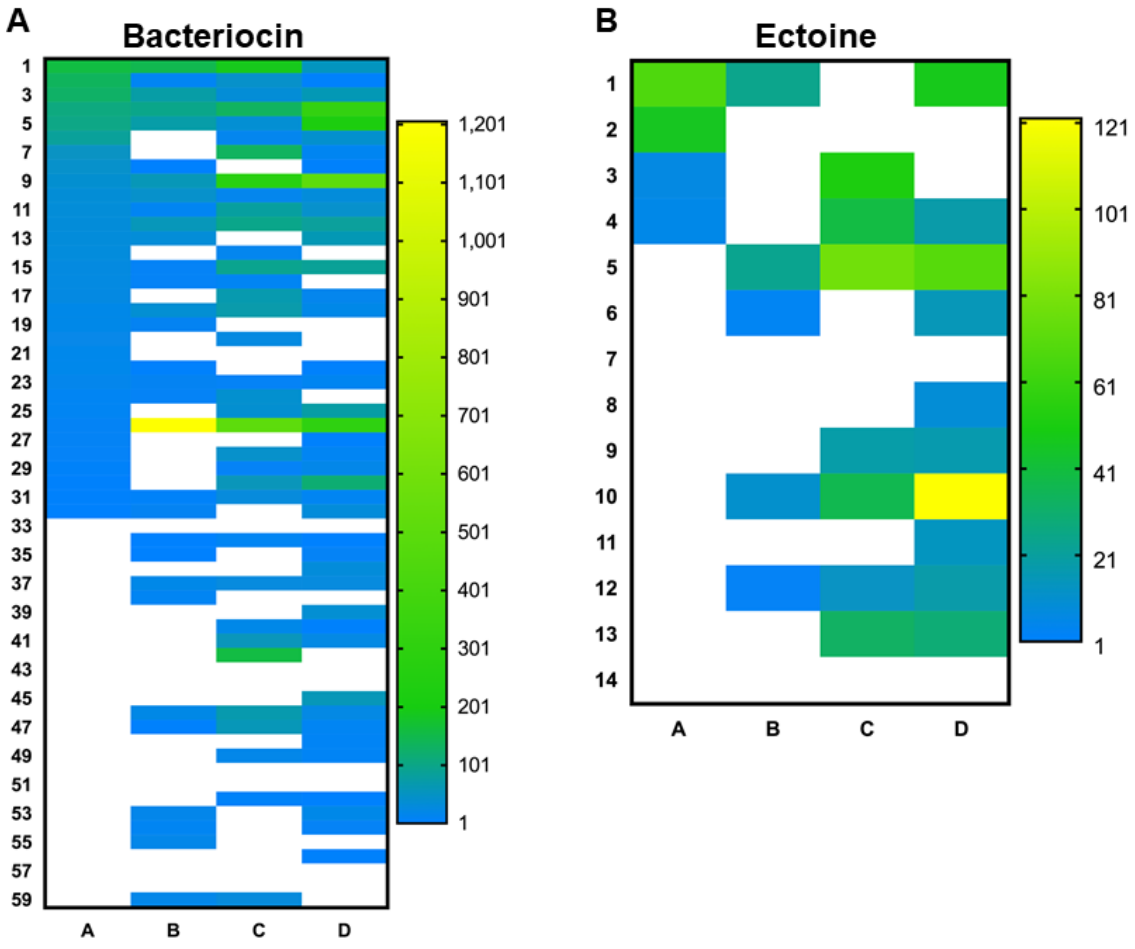


Figure 2. Heatmaps of bacteriocin and ectoine core biosynthetic genes expressed in ANG/JC metatranscriptomes. Expression of core **(A)** bacteriocin and **(B)** ectoine genes. Expression is calculated as the average normalized TPM values per sample. Row numbers represent each core biosynthetic genes for all predicted clusters. White cells are genes with TPM expression less than one. Column letters correspond to metatranscriptome sample types: A = ANG, B = day 0 JCs, C = day 10 control JCs, D = day 10 challenge JCs.

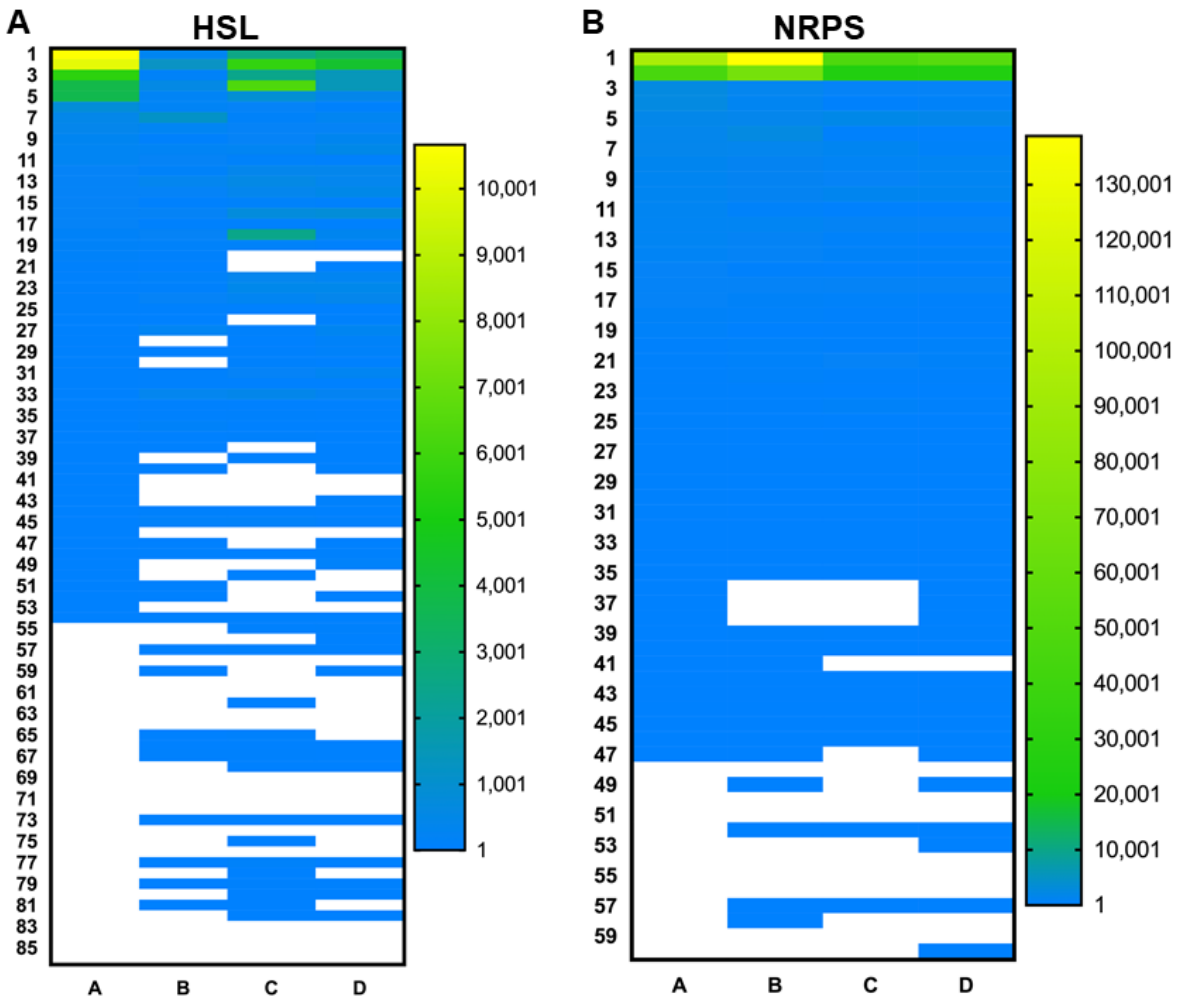


Figure 3. Heatmaps of HSL and NRPS core biosynthetic genes expressed in ANG/JC metatranscriptomes. Expression of core **(A)** homoserine lactone (HSL) and **(B)** nonribosomal peptide synthetase (NRPS) genes. Expression is calculated as the average normalized TPM values per sample. Row numbers represent each core biosynthetic genes for all predicted clusters. White cells are genes with TPM expression less than one. Column letters correspond to metatranscriptome sample types: A = ANG, B = day 0 JCs, C = day 10 control JCs, D = day 10 challenge JCs.

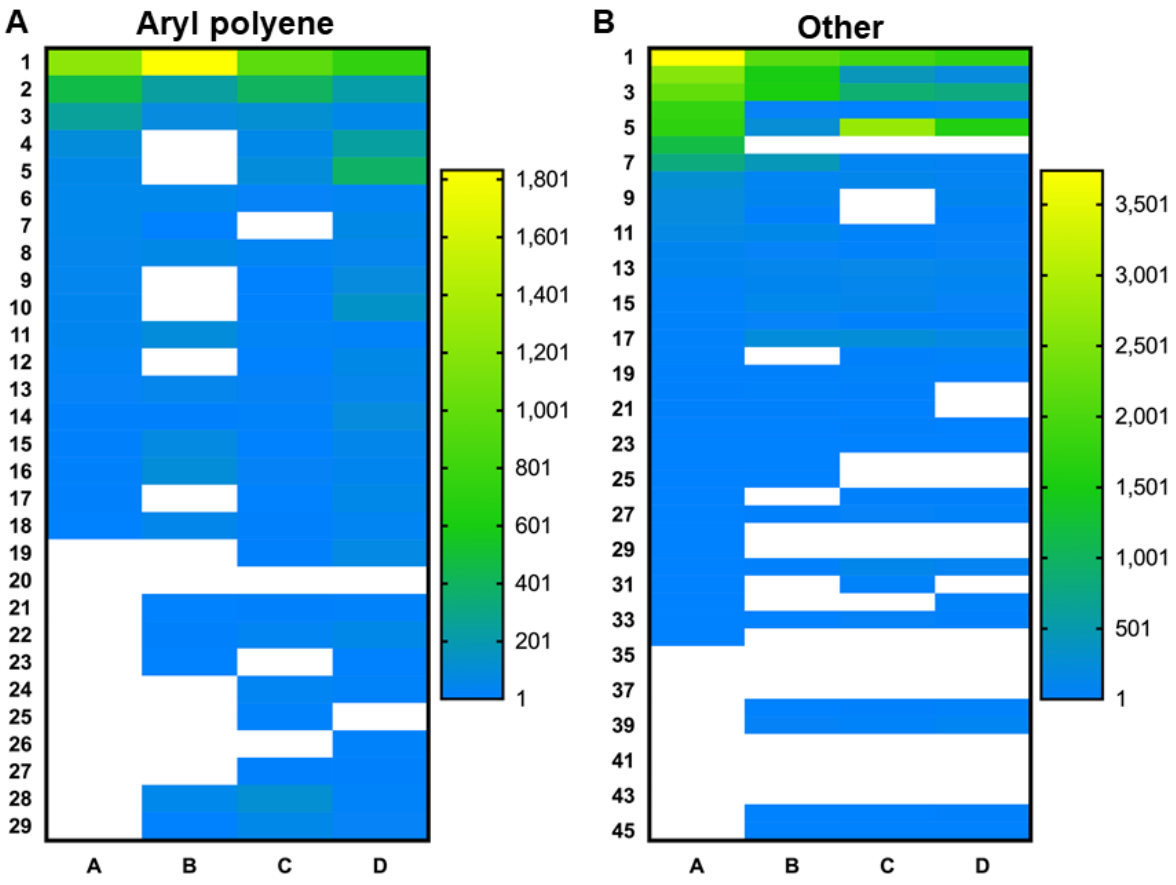


Figure 4. Heatmaps of aryl polyene and other core biosynthetic genes expressed in ANG/JC metatranscriptomes. Expression of core **(A)** aryl polyene and **(B)** other genes. Expression is calculated as the average normalized TPM values per sample. Row numbers represent each core biosynthetic genes for all predicted clusters. White cells are genes with TPM expression less than one. Column letters correspond to metatranscriptome sample types: A = ANG, B = day 0 JCs, C = day 10 control JCs, D = day 10 challenge JCs.

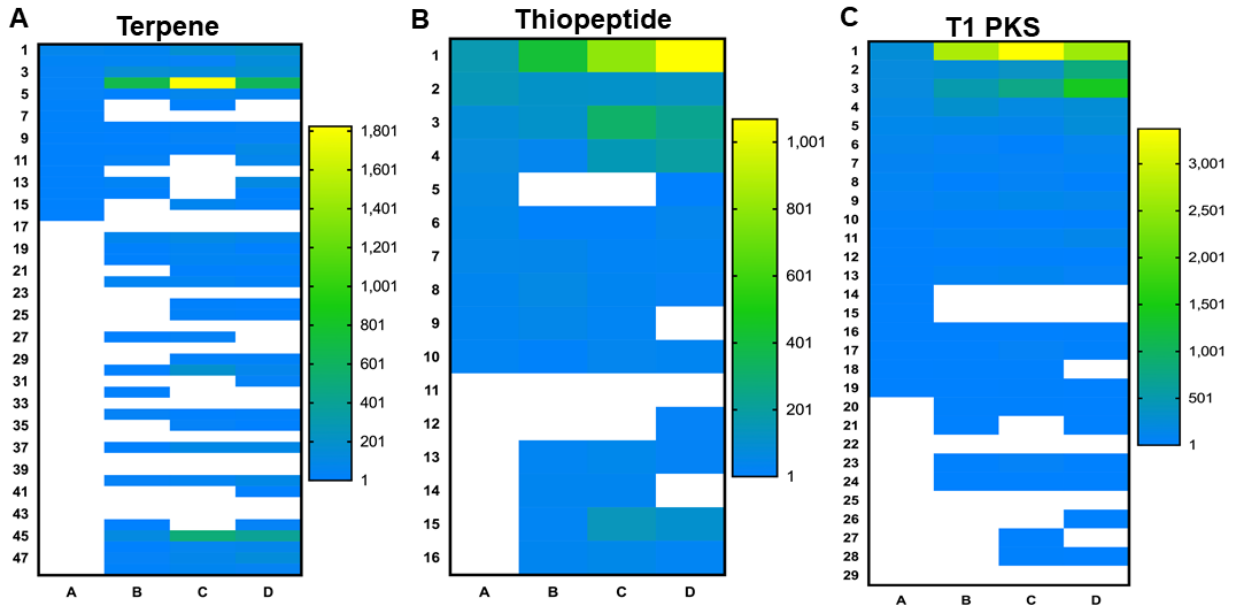


Figure 5. Heatmaps of terpene, thiopeptide, and T1 PKS core biosynthetic genes expressed in ANG/JC metatranscriptomes. Expression of core **(A)** terpene, **(B)** thiopeptide, and **(C)** type 1 polyketide synthase (T1 PKS) genes. Expression is calculated as the average normalized TPM values per sample. Row numbers represent each core biosynthetic genes for all predicted clusters. White cells are genes with TPM expression less than one. Column letters correspond to metatranscriptome sample types: A = ANG, B = day 0 JCs, C = day 10 control JCs, D = day 10 challenge JCs.

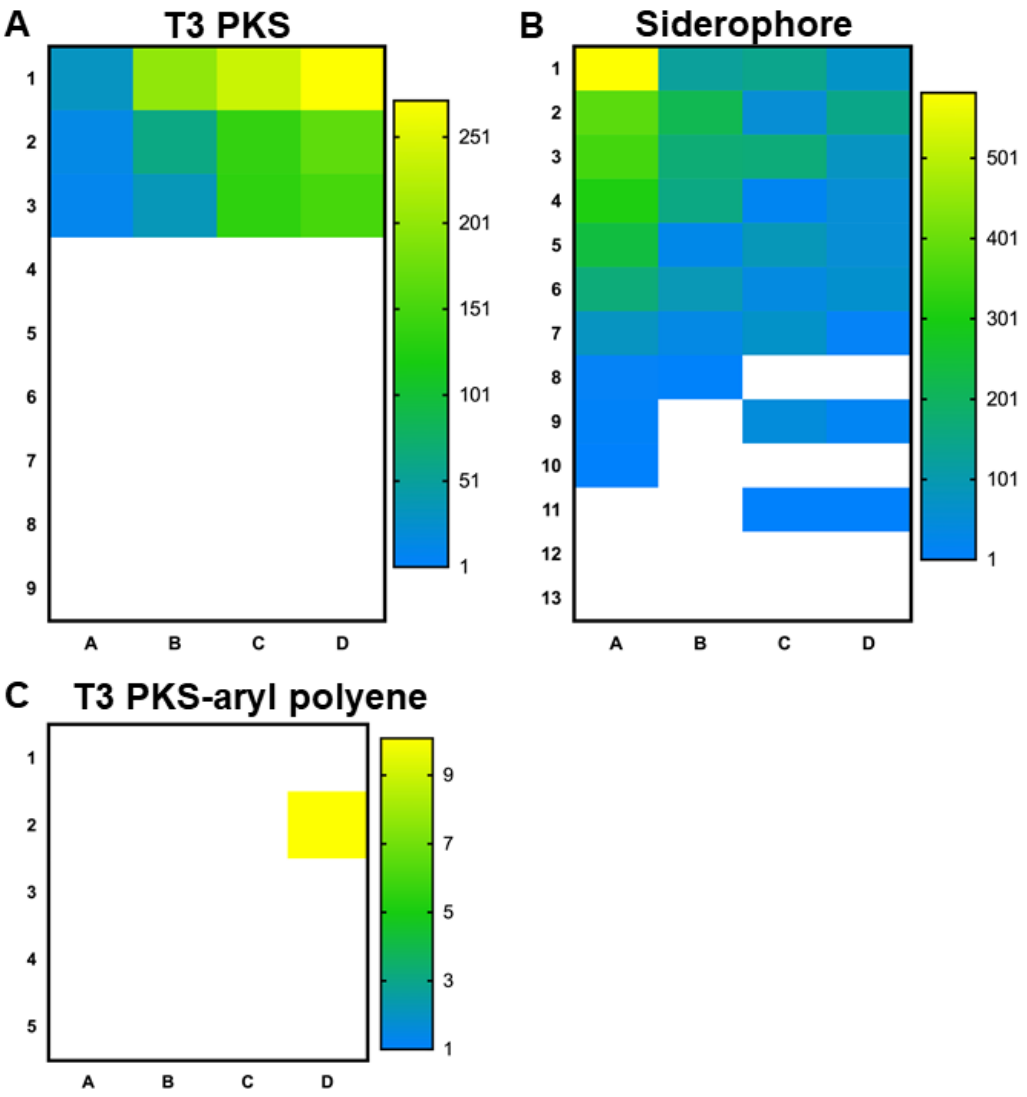


Figure 6. Heatmaps of T3 PKS, siderophore, and T3 PKS-aryl polyene core biosynthetic genes expressed in ANG/JC metatranscriptomes. Expression of core **(A)** type 3 polyketide synthase (T3 PKS), **(B)** siderophore, and **(C)** T3 PKS-aryl polyene genes. Expression is calculated as the average normalized TPM values per sample. Row numbers represent each core biosynthetic genes for all predicted clusters. White cells are genes with TPM expression less than one. Column letters correspond to metatranscriptome sample types: A = ANG, B = day 0 JCs, C = day 10 control JCs, D = day 10 challenge JCs.

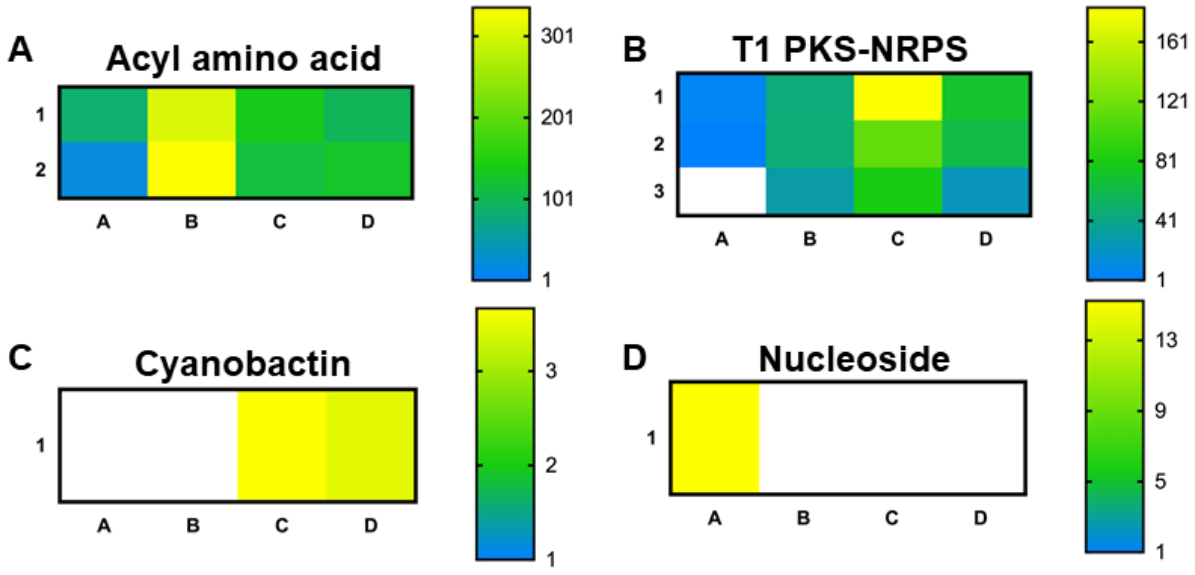


Figure 7. Heatmaps of acyl amino acid, T1 PKS-NRPS, cyanobactin, and nucleoside core biosynthetic genes expressed in ANG/JC metatranscriptomes. Expression of core **(A)** acyl amino acid, **(B)** type 1 polyketide synthase (T1 PKS)-non ribosomal peptide synthetase (NRPS), **(C)** cyanobactin, and **(D)** nucleoside genes. Expression is calculated as the average normalized TPM values per sample. Row numbers represent each core biosynthetic genes for all predicted clusters. White cells are genes with TPM expression less than one. Column letters correspond to metatranscriptome sample types: A = ANG, B = day 0 JCs, C = day 10 control JCs, D = day 10 challenge JCs.

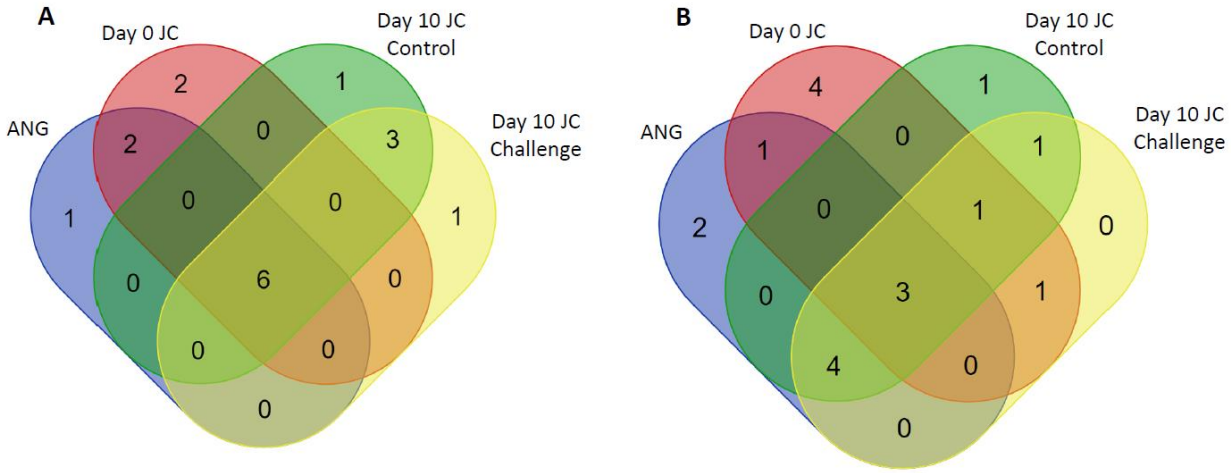


Figure 8. Overlap of the top 10 highest expressed secondary metabolite genes between the ANG and JC metatranscriptomes. (A) Number of all biosynthetic and neighbor genes in a cluster highly expressed between the ANG/JC samples. Shared genes are listed in Table S6. **(B)** Number of highest expressed core biosynthetic genes shared between the ANG/JC samples. Shared genes are listed in Table S7.

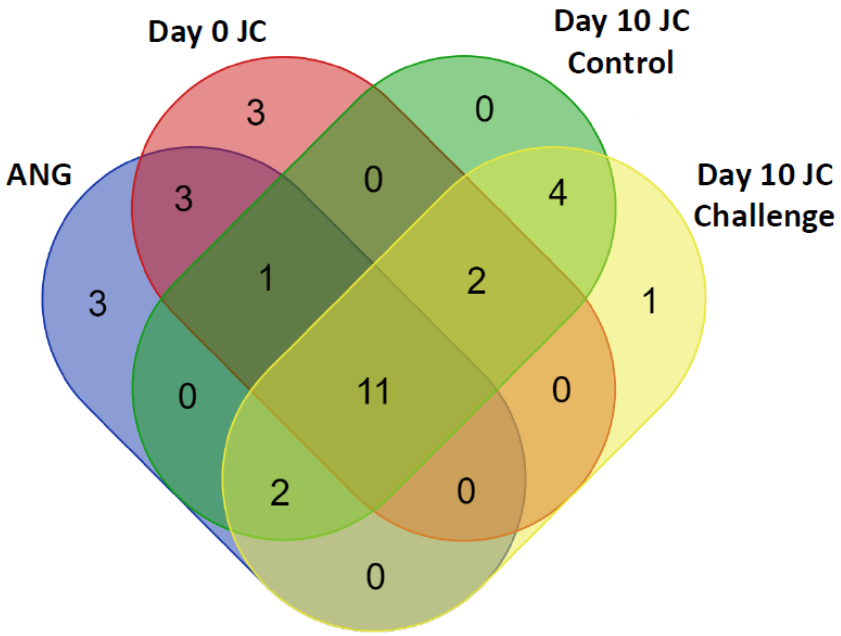


Figure 9. Overlap of the top 20 highest expressed genes in the *Verrucomicrobia* PD13 Bin 1 MAG. Shared genes are listed in Table S10.

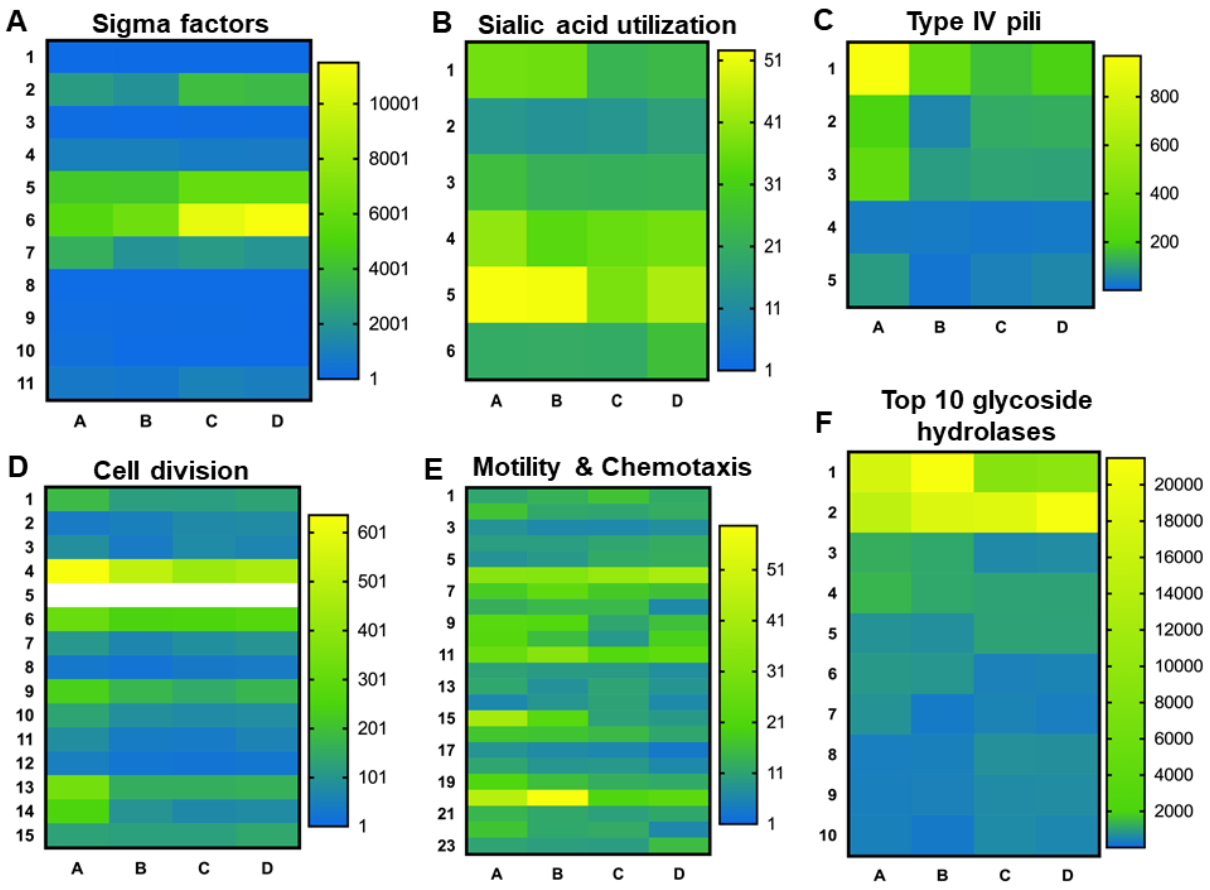


Figure 10. Heatmaps of *Verrucomicrobia* PD13 Bin 1 genes expressed in ANG/JC metatranscriptomes. Expression of **(A)** sigma factor, **(B)** sialic acid utilization, **(C)** type IV pili, **(D)** cell division, **(E)** motility and chemotaxis, and **(F)** the top 10 highest expressed glycoside hydrolase genes. Expression is calculated as the average normalized TPM values per sample. Row numbers correspond to genes in Appendix V. White cells are genes with TPM expression less than one. Column letters correspond to metatranscriptome sample types: A = ANG, B = day 0 JCs, C = day 10 control JCs, D = day 10 challenge JCs.

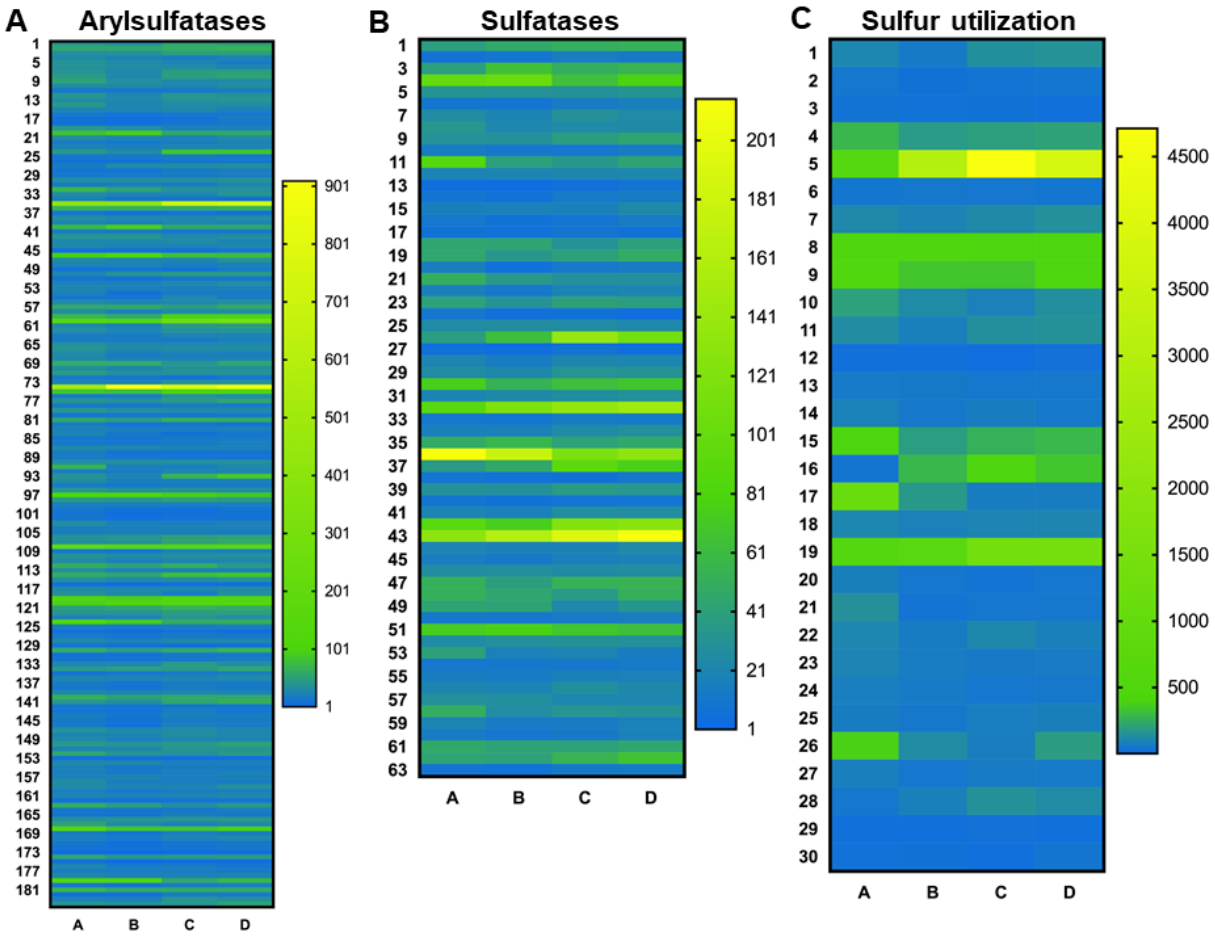


Figure 11. Heatmaps of *Verrucomicrobia* PD13 Bin 1 genes expressed in ANG/JC metatranscriptomes. Expression of **(A)** arylsulfatase, **(B)** sulfatase, and **(C)** other sulfur utilization genes. Expression is calculated as the average normalized TPM values per sample. Row numbers correspond to genes in Appendix V. Column letters correspond to metatranscriptome sample types: A = ANG, B = day 0 JCs, C = day 10 control JCs, D = day 10 challenge JCs.

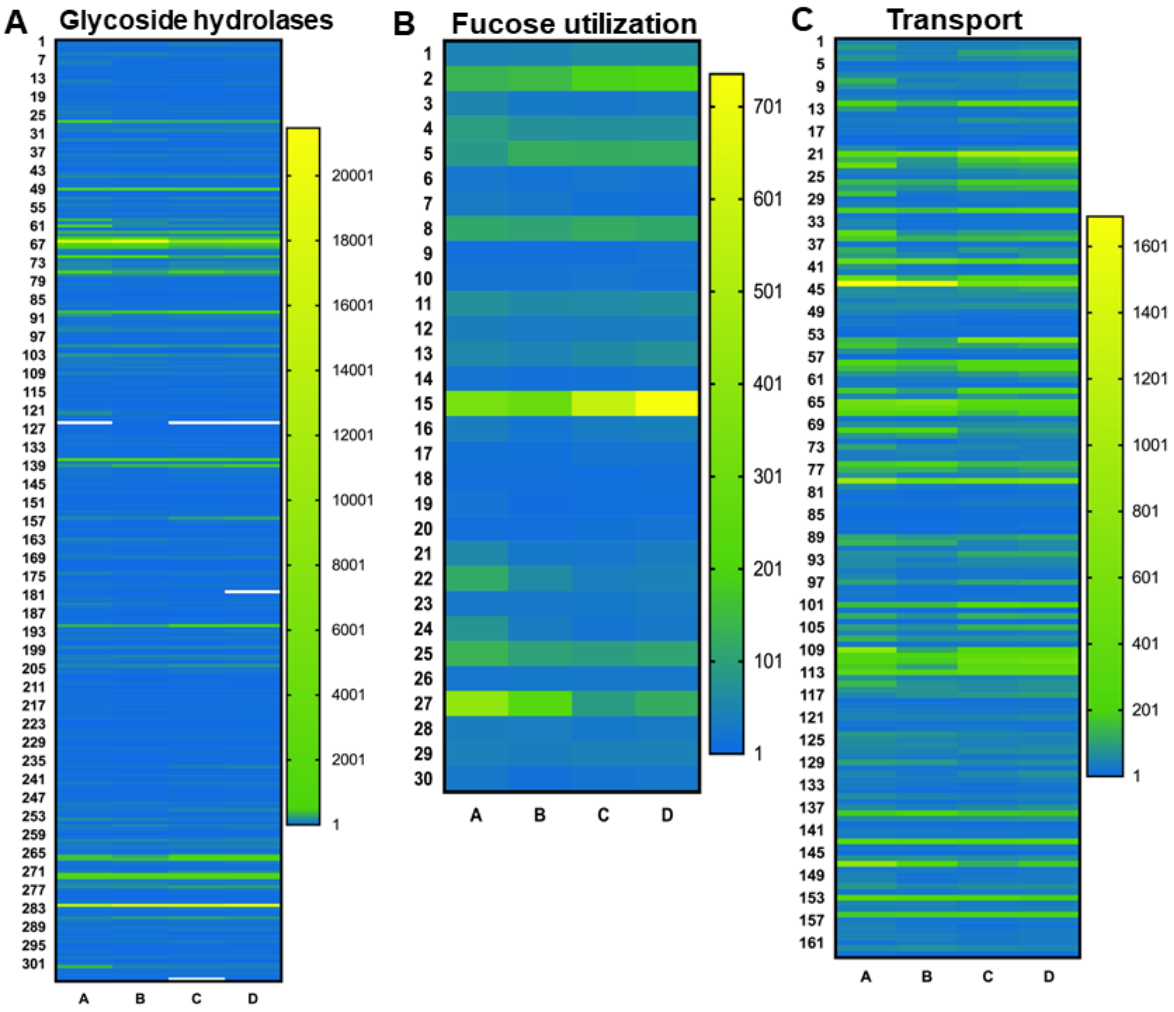


Figure 12. Heatmaps of *Verrucomicrobia* PD13 Bin 1 genes expressed in ANG/JC metatranscriptomes. Expression of **(A)** all glycoside hydrolase, **(B)** fucose utilization, and **(C)** transport genes. Expression is calculated as the average normalized TPM values per sample. Row numbers correspond to genes in Appendix V. White cells are genes with TPM expression less than one. Column letters correspond to metatranscriptome sample types: A = ANG, B = day 0 JCs, C = day 10 control JCs, D = day 10 challenge JCs.

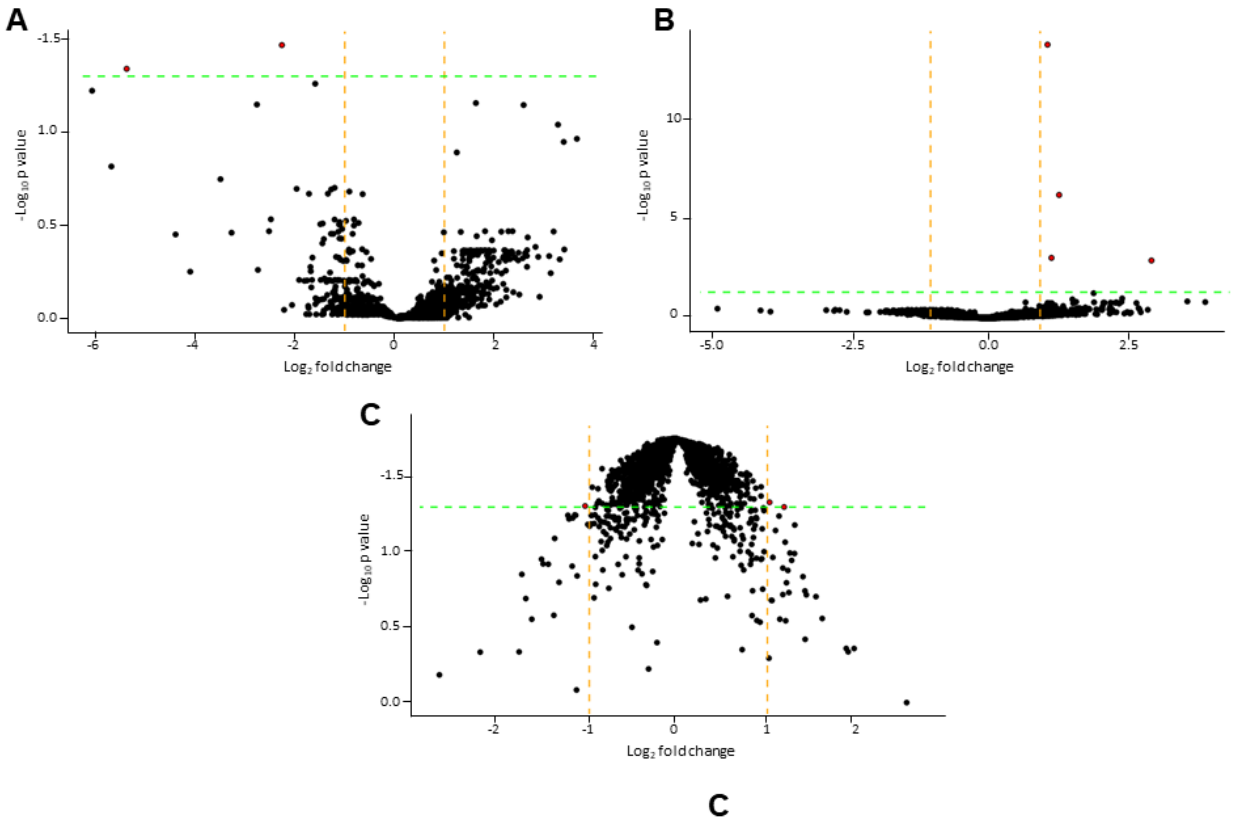


Figure 13. Volcano plots of *Verrucomicrobia* PD13 Bin 1 mapped metatranscriptome reads. Comparisons of **(A)** ANG versus day 0 JCs, **(B)** day 0 JCs versus day 10 control JCs, and **(C)** day 10 challenge versus day 10 control JCs. Green dotted lines indicate *p* value cut off of 0.05 (-log10 *p* value = 1.3). Dotted orange lines indicate log fold change cut off of 2 (log_2 fold change = 1).

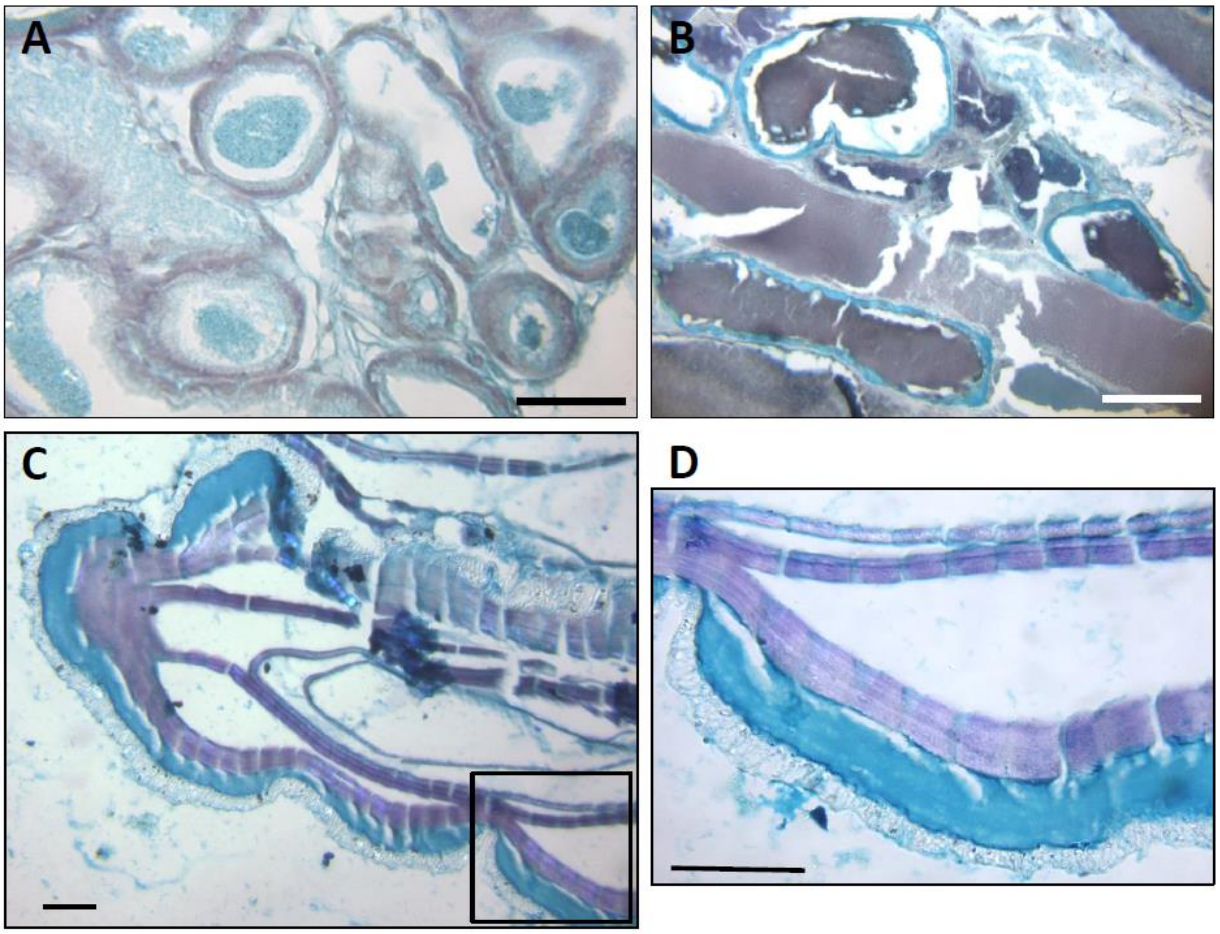


Figure 14. Staining of mucins in ANG tubules and egg JCs. (A) The interior of some tubules stained blue for sialic acid-containing mucins while the epithelia stained purple for sulfate-containing mucins. **(B)** In another portion of the same ANG section, an opposite staining pattern was observed. **(C)** Sulfomucins and sialomucins were observed in distinct layers of the egg jelly coat. **(D)** A magnified image of the jelly coat layers in boxed inset (B). Scale bars = 60 μm.

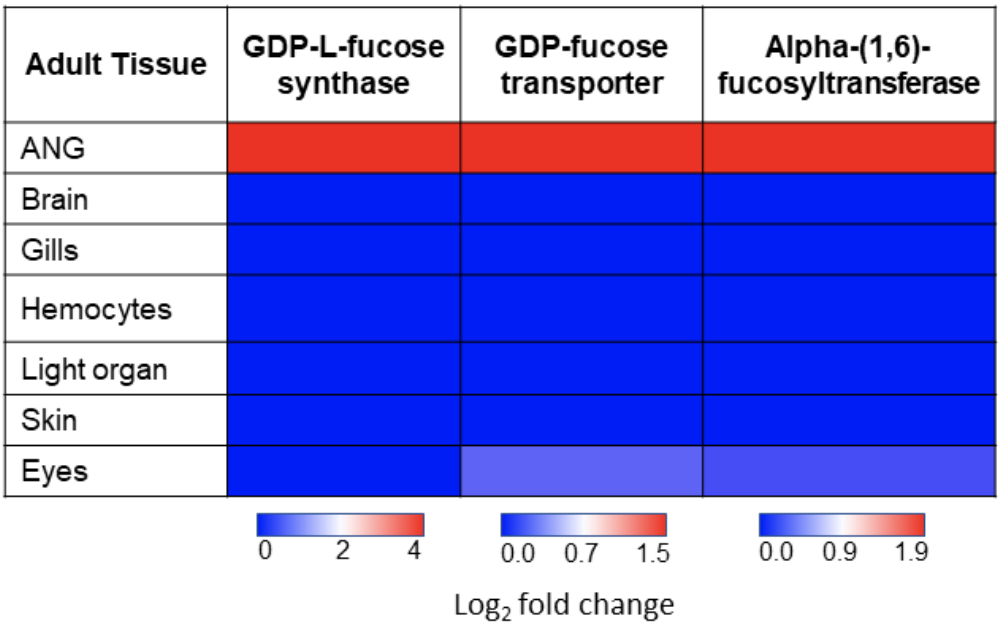


Figure 15. Expression of genes for the synthesis of fucose in adult *E. scolopes* tissues.

Expression of host genes for the synthesis of fucose (GDP-L-fucose synthase), transfer of fucose to the golgi complex (GDP-fucose transporter), and attachment to glycopeptides (alpha-1,6-fucosyltransferase) is shown in different adult tissues. Log_2 fold change in expression is shown for each gene, with red indicating higher expression and blue indicating low expression.

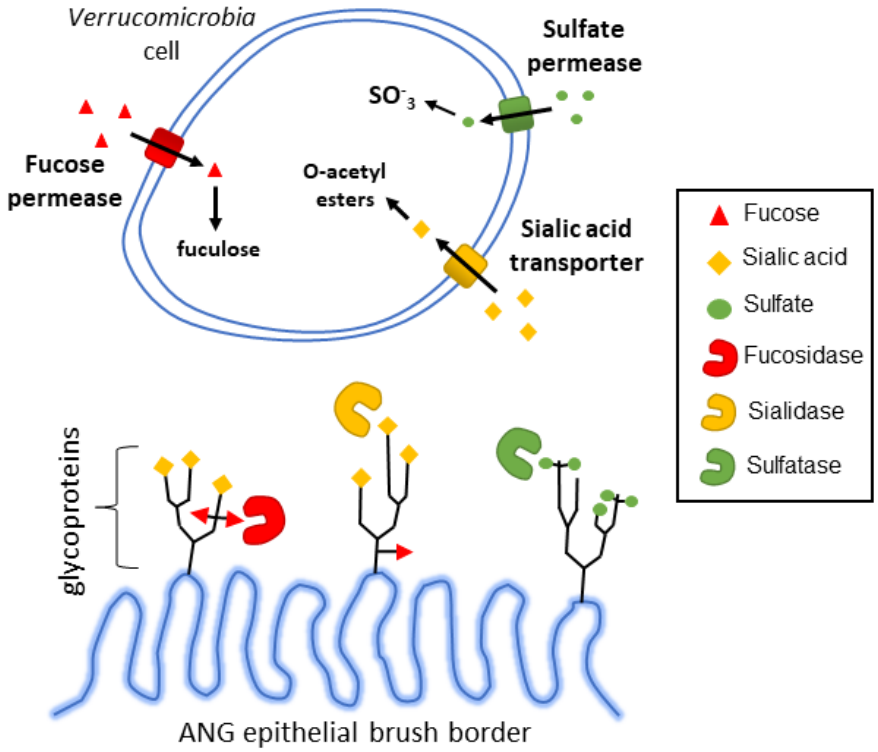


Figure 16. Predicted model of glycoprotein usage in the ANG by *Verrucomicrobia* PD13 Bin1. Fucosidases, sialidases, and sulfatases are expressed by the *Verrucomicrobia* PD13 Bin 1 MAG in the ANG, and can potentially cleave fucose, sialic acid, and sulfate from host-produced glycoproteins. Permeases or transporters for each of these nutrients are expressed by *Verrucomicrobia* PD13 Bin 1 in the ANG, as well as genes to further utilize these compounds in the production of fuculose, o-acetyl esters, and sulfite.

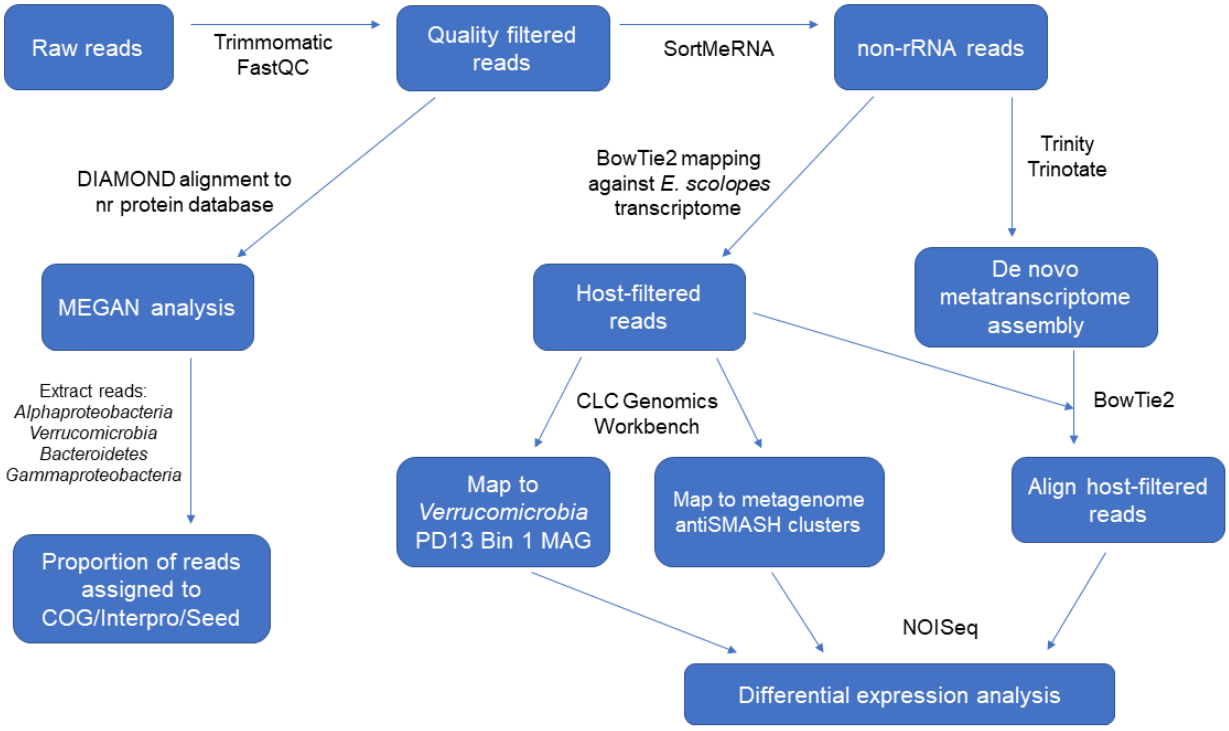


Figure S1. Flow chart of metatranscriptome analyses. Bioinformatic tools used and notes appear in black next to connecting lines.

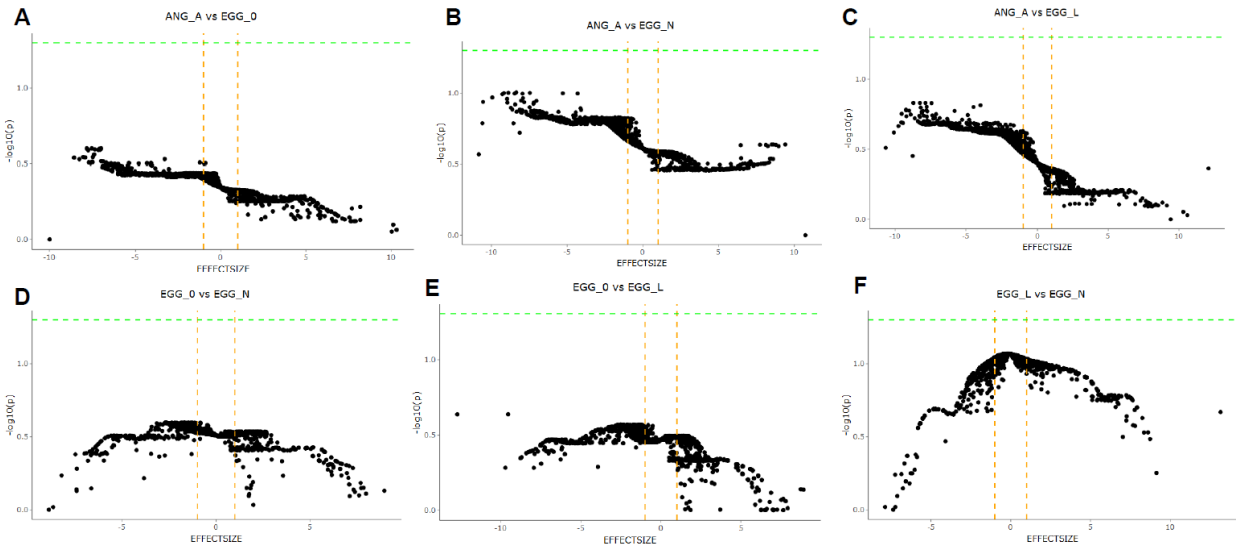


Figure S2. Volcano plots of antiSMASH mapped metatranscriptome reads. Comparisons of (A) ANG versus day 0 JCs, (B) ANGs versus day 10 control JCs, (C) ANGs versus day 10 challenge JCs, (D) day 0 JCs versus day 10 control JCs, (E) day 0 JCs versus day 10 challenge JCs, and (F) day 10 challenge versus day 10 control JCs. Green dotted lines indicate *p* value cut off of 0.05 (-log10 *p* value = 1.3). Dotted orange lines indicate log fold change cut off of 2 (log_2 fold change/ "EFFECTSIZE" = 1).

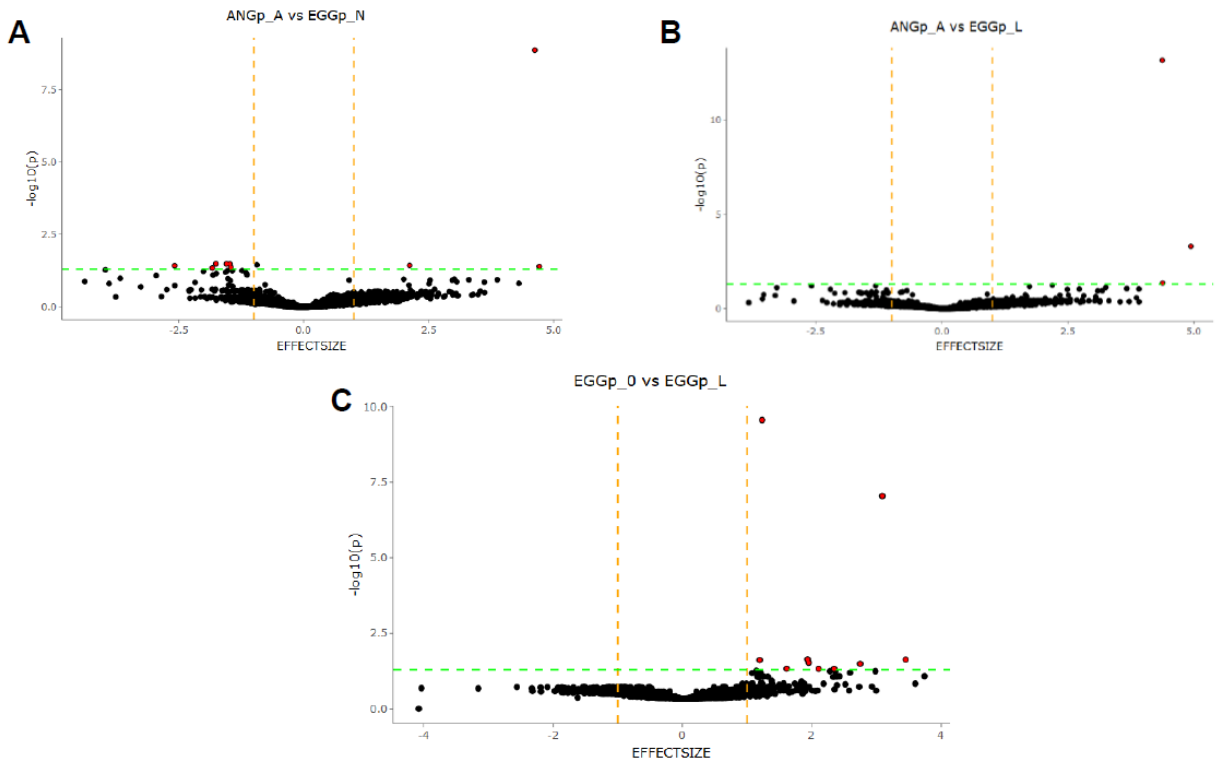


Figure S3. Other volcano plots of *Verrucomicrobia* PD13 Bin1 mapped metatranscriptome read sample comparisons. Comparisons of (A) ANGs versus day 10 control JCs, (B) ANGs versus day 10 challenge JCs, and (C) day 0 JCs versus day 10 challenge JCs. Green dotted lines indicate *p* value cut off of 0.05 (-log10 *p* value = 1.3). Dotted orange lines indicate log fold change cut off of 2 (log_2 fold change/ "EFFECTSIZE" = 1).

Table S1. ANG Metatranscriptome Samples

Female	Mantle length (mm)	RNA concentration (ng/µl)	RNA integrity (RIN) number	Eukaryotic mRNA selection kit input (ng)	Eukaryotic mRNA Qubit concentration (ng/µl)	Prokaryotic RNA Qubit concentration (ng/µl)
E8	25	38.7	7.1	1,536	<0.02	31.6
E16	25	29.7	NA	918.4	<0.02	25.6
J1	27	416	7.4	4,160	<0.02	26.9
J8	24	217	7.7	2,170	<0.02	26.9
K3	26	387	6.8	3,870	<0.02	23.2
K7	23	399	6	3,990	<0.02	47.3

Table S2. Egg Jelly Coat Metatranscriptome Samples

Female	Clutch number	Clutch age (day)	Condition	Number of pooled JCs	RNA concentration (ng/µl)	RNA integrety (RIN) number
E16	18	0	NA	30	<0.02	2.1
E16	39	10	Control	30	<0.02	5.3
E16	39	10	Challenge	30	<0.02	5
E8	32	0	NA	30	2	3
E8	35	10	Control	30	<0.02	3.7
E8	35	10	Challenge	30	<0.02	4.6
J1	6	0	NA	40	<0.02	3.7
J1	13	10	Control	40	<0.02	NA
J1	13	10	Challenge	30	<0.02	2.6
J8	7	0	NA	40	<0.02	3.6
J8	14	10	Control	40	<0.02	3.5
J8	14	10	Challenge	30	<0.02	2.9
K3	10	0	NA	40	<0.02	3.3
K3	15	10	Control	40	<0.02	4
K3	15	10	Challenge	40	<0.02	4.1
K7	12	0	NA	40	<0.02	4.5

Table S3. Amount of raw, trimmed, and rRNA metatranscriptomic reads

Sample type	Female	Raw read pairs	Trimmed read pairs	non-rRNA read pairs	% non-rRNA read pairs	rRNA read pairs	% rRNA read pairs
ANG (prokaryotic)	E8	54355111	49971802	38912432	77.9	11059370	22.1
	E16	50004451	46276567	43193917	93.3	3082650	6.66
	J1	25367858	24078275	21886040.5	90.9	2192234.5	9.1
	J8	26163304	24366336	20148940	82.7	4217396	17.3
	K3	25524305	23869970	23231797.5	97.3	638172.5	2.67
	K7	28299703	26632148	20866262.5	78.4	5765885.5	21.7
ANG (eukayrotic)	E8	45018071	40586214	39878502.5	98.3	707711.5	0.98
	E16	50890806	47207382	46482960	98.5	724422	1.53
	J1	30350048	28396586	28081298	98.9	315288	1.11
	J8	26856936	25327460	25076031.5	99.0	251428.5	0.99
	K3	25654054	24365387	24097223	98.9	268164	1.1
	K7	22597439	20905108	20637094	98.7	268014	1.28
Day 0 JC	E8	35740484	32898068	7659898	23.3	25238170	76.7
	E16	60698982	57390391	36397979.5	63.4	20992411.5	36.6
	J1	17422209	16572377	6540413.5	39.5	10031963.5	60.5
	J8	28411406	26920765	9696258	36.0	17224507	64.0
	K3	25439208	23349623	11532659.5	49.4	11816963.5	50.6
	K7	23767794	22073985	7368796.5	33.4	14705188.5	66.6
Day 10 JC control	E8	35466550	33677374	10656361.5	31.6	23021012.5	68.4
	E16	40760430	38767011	10305677	26.6	28461334	73.4
	J1	22241939	20849392	10620570	50.9	10228822	49.1
	J8	30526502	28619682	9658557	33.8	18961125	66.3
	K3	20642074	19138978	6925195.5	36.2	12213782.5	63.8
Day 10 JC challenge	E8	41421128	39493207	7466418.5	18.9	32026788.5	81.1
	E16	72703859	72703859	15971933	22.0	56731926	78.0
	J1	24177477	22571494	9561764	42.4	13009730	57.6
	J8	26773970	25569703	9778813.5	38.2	15790889.5	61.8
	K3	21498046	19992736	7658534	38.3	12334202	61.7

Table S4. Percentages of rRNA metatranscriptome reads

Sample type	Female	% bac. 16S reads	% bac. 23S reads	% arc. 16S reads	% arch. 23S reads	% euk. 18S reads	% euk. 28S reads	% rfam 5S reads	% rfam 5.8s reads
ANG (prokaryotic)	E8	0.9	1.03	0.47	1.93	10.8	5.84	0.54	0.6
	E16	0.62	0.7	0.07	0.55	2.39	1.9	0.21	0.23
	J1	1.57	4.33	0.04	0.47	1.06	1.54	0.02	0.07
	J8	1.33	3.83	0.19	1.93	4.98	4.68	0.06	0.31
	K3	0.24	0.5	0.01	0.21	0.64	0.95	0.04	0.09
	K7	1.73	4.19	0.34	1.83	8.96	4.25	0.07	0.29
ANG (eukayrotic)	E8	0.98	0.2	0	0.02	0.2	0.3	0.04	0.01
	E16	0.59	0.27	0	0.01	0.25	0.37	0.03	0.01
	J1	0.26	0.23	0	0.02	0.27	0.28	0.05	0
	J8	0.24	0.15	0	0.01	0.19	0.37	0.03	0
	K3	0.27	0.23	0	0	0.23	0.34	0.04	0
	K7	0.41	0.25	0	0	0.26	0.31	0.04	0
Day 0 JC	E8	27.2	49.2	0.02	0.05	0.16	0.1	0.01	0
	E16	14.7	21.6	0.01	0.02	0.08	0.06	0.18	0
	J1	22.2	38.3	0	0.01	0.03	0.03	0.02	0
	J8	19.7	44.1	0.01	0.02	0.08	0.05	0.08	0
	K3	17.0	33.3	0.01	0.03	0.11	0.07	0.07	0
	K7	26.3	40.0	0.01	0.03	0.14	0.07	0.06	0
Day 10 JC control	E8	26.5	39.3	0.1	0.29	1.04	1.14	0.02	0
	E16	28.1	44.5	0.03	0.08	0.26	0.3	0.16	0
	J1	17.3	30.6	0.04	0.12	0.46	0.57	0.02	0
	J8	20.4	42.4	0.07	0.28	1.42	1.63	0.05	0
	K3	20.9	38.5	0.12	0.43	1.73	2.08	0.05	0
Day 10 JC challenge	E8	32.7	45.0	0.18	0.47	1.33	1.36	0.11	0
	E16	30.4	43.0	0.24	0.51	1.81	2.04	0.04	0
	J1	19.3	37.5	0.03	0.13	0.32	0.41	0.01	0
	J8	18.1	38.7	0.17	0.52	1.85	2.33	0.04	0
	K3	20.8	38.2	0.09	0.26	1.06	1.3	0.04	0

Abbreviations: bac. = bacterial, arc. = archaeal, euk. = eukaryotic

Table S5. Number of metatranscriptome reads that aligned to the *Euprymna scolopes* transcriptome

Sample type	Female	Number of read pairs that aligned to *E. scolopes* transcriptome	Percent *E. scolopes* read pairs	Number of read pairs that did not align to *E. scolopes* transcriptome	Percent non-*E. scolopes* read pairs
ANG (prokaryotic)	E8	27324169	54.1	27030942	49.7
	E16	13507889	29.1	36496562	73.0
	J1	7463235	31.8	17904623	70.6
	J8	11790965	48.1	14372339	54.9
	K3	6693060	28.1	18831245	73.8
	K7	14158863	52.4	14140840	50.0
Day 0 JC	E8	17944156	55.4	17796328	49.8
	E16	12444474	22.9	48254508	79.5
	J1	8406711	51.8	9015498	51.7
	J8	14076261	53.2	14335145	50.5
	K3	8241143	35.8	17198065	67.6
	K7	9463288	43.6	14304506	60.2
Day 10 JC control	E8	9702226	31.5	25764324	72.6
	E16	9904643	28.2	30855787	75.7
	J1	8487459	41.4	13754480	61.8
	J8	13093222	46.1	17433280	57.1
	K3	6631395	35.8	14010679	67.9
Day 10 JC challenge	E8	12184360	33.6	29236768	70.6
	E16	18169964	27.5	58259939	76.2
	J1	9920041	44.8	14257436	59.0
	J8	11009833	43.9	15764137	58.9
	K3	6672987	34.5	14825059	69.0

Table S6. Secondary metabolite genes with highest expression in ANG and JC metatranscriptomes. Expression values are TPM normalized with standard deviation.

Samples	Number of shared genes	Cluster number	Cluster type	Cluster gene	Gene type	Average expression (TPM)			
						ANG	Day 0 JC	Day 10 JC Control	Day 10 JC Challenge
ANG, Day 0 JC, Day 10 JC challenge, Day 10 JC Control	6	220	NRPS	22	bio	94,835.7 ± 62,115.4	138,747.0 ± 62,770.5	48,035.9 ± 30,110.6	52,857.6 ± 37,152.1
		77	NRPS	8	other	46,559.3 ± 50,880.1	50,291.0 ± 26,836.2	69,474.9 ± 77,217.7	67,491.8 ± 84,615.6
		77	NRPS	7	other	96,706.4 ± 97,786.4	106,642.8 ± 68,152.4	103,907.4 ± 67,436.3	88,613.7 ± 60,077.1
		220	NRPS	30	other	50,217.7 ± 34,827.3	64,980.5 ± 52,657.3	42,140.9 ± 43,254.2	54,193.8 ± 64,965.1
		194	NRPS	15	bio	44,902.7 ± 30,862.4	68,050.1 ± 30,509.3	25,337.9 ± 18,855.9	25,337.0 ± 20,490.5
		229	NRPS	16	other	13,907.8 ± 13,773.75	14,263.7 ± 6,172.4	15,568.4 ± 11,699.7	14,026.3 ± 12,283.4
ANG, Day 0 JC	2	207	other	4	other	13,403.10 ± 14,673.0	11,272.2 ± 6,668.1	7,257.2 ± 6,209.0	7,218.1 ± 7,762.0
		194	NRPS	24	other	24,589.3 ± 23,553.9	19,769.3 ± 14,443.2	9,072.6 ± 11,141.1	11,223.7 ± 14,025.7
Day 10 JC Challenge, Day 10 JC Control	3	82	Bacteriocin	1	other	19,544.4 ± 43,290.1	1,279.9 ± 2,055.7	1,710.2 ± 3,852.6	11,236.8 ± 24,560.7
		253	Bacteriocin	574	other	12,361.9 ± 6,359.4	656.2 ± 641.7	5,033.1 ± 1,869.1	16,220.2 ± 11,184.9
		25	terpene	25	other	17,019.8 ± 10,522.8	6,009.0 ± 6,526.7	5,993.2 ± 4,738.7	22,184.9 ± 16,257.4
ANG	1	51	other	63	other	34,761.7 ± 77,745.1	1,687.4 ± 3,405.2	4,586.7 ± 6,111.5	2,692.4 ± 3,226.7

D0 JC	2	194	NRPS	16	other	20,601.7 ± 11,591.7	10,876.1 ± 5,956.0	7,774.0 ± 5,631.5	7,440.4 ± 5,891.9
		156	NRPS	23	other	23,804.6 ± 58,309.0	2,550.9 ± 6,248.5	9,981.5 ± 22,319.3	6,938.9 ± 15,515.8
Day 10 JC Control	1	192	HSL	18	other	12,462.6 ± 5,491.0	4,974.0 ± 3,487.2	7,205.7 ± 5,283.1	8,368.6 ± 4,393.7
Day 10 JC Challenge	1	104	arylpolyene	38	other	13,102.6 ± 23,926.1	253.5 ± 447.0	1,167.7 ± 1,701.3	788.1 ± 1,080.6

Table S7. Highest expressed biosynthetic secondary metabolite genes in ANG and JC metatranscriptomes. Expression values are TPM normalized with standard deviation.

Samples	Number of shared genes	Cluster number	Cluster type	Cluster gene	Average expression (TPM)			
					ANG	Day 0 JC	Day 10 JC Control	Day 10 JC Challenge
ANG, Day 0 JC, Day 10 JC Challenge, Day 10 JC Control	3	220	NRPS	22	94,835.7 ± 62,115.4	138,747.0 ± 62,770.5	4,8035.9 ± 30,110.6	52,857.6 ± 37,152.1
		194	NRPS	15	44,902.7 ± 30,862.4	68,050.0 ± 30,509.3	25,337.9 ± 18,855.9	25,337.0 ± 20,490.5
		207	other	1	3,741.0 ± 3,899.7	2,141.3 ± 1,653.0	1,933.6 ± 2,428.2	1,750.2 ± 2,244.9
ANG, Day 10 JC Challenge, Day 10 JC Control	4	43	HSL	2	10,043.2 ± 9,200.0	1,223.3 ± 1,239.5	5,686.2 ± 7,534.5	4,373.8 ± 6,819.9
		11	HSL	3	5,486.5 ± 12,924.1	129.1 ± 316.2	2,462.1 ± 5,236.1	1,466.2 ± 3,189.8
		144	HSL	39	10,665.9 ± 10,869.2	293.2 ± 533.3	2,524.9 ± 2,928.4	3,253.3 ± 5,218.3
		32	HSL	8	3,796.6 ± 6,189.4	569.3 ± 1,300.9	6,320.2 ± 13,383.6	1,491.8 ± 2,927.8
Day 0 JC, Day 10 JC Challenge, Day 10 JC Control	1	120	T1PKS	20	255.7 ± 284.7	2,704.8 ± 4,964.1	3,373.3 ± 6,529.5	2,582.3 ± 5,037.5
ANG, Day 0 JC	1	124	other	5	2,575.4 ± 3,157.5	1,483.3 ± 2,191.1	417.2 ± 489.5	215.0 ± 222.2
Day 0 JC, Day 10 JC Challenge	1	113	NRPS	4	1,792.0 ± 2,531.8	1,566.0 ± 2,163.4	1,907.4 ± 2,546.2	1,702.0 ± 2,642.7

Day 10 JC Challenge, Day 10 JC Control	1	119	other	2	1,719.8 ± 2,223.7	266.4 ± 479.6	2,732.1 ± 6,020.2	1,595.3 ± 3,389.3
ANG	2	125	NRPS	1	2,540.0 ± 4,289.3	1,332.3 ± 977.7	536.6 ± 417.8	596.6 ± 520.2
		190	HSL	6	3,680.6 ± 7,758.2	201.1 ± 255.3	933.8 ± 1,261.4	424.7 ± 528.2
Day 0 JC	4	220	NRPS	27	1,541.3 ± 1,324.8	1,427.6 ± 1,654.3	1,176.0 ± 1,604.8	497.4 ± 781.3
		194	NRPS	21	1,568.6 ± 1,652.8	2,489.5 ± 4,802.1	234.8 ± 177.6	304.9 ± 443.5
		200	other	2	2,222.5 ± 2,434.1	1,512.0 ± 699.0	890.8 ± 737.2	804.4 ± 807.3
		211	arylpolyene	4	1,222.8 ± 1,613.4	1,833.4 ± 1,925.9	959.7 ± 895.7	734.3 ± 563.8
Day 10 JC Control	1	54	HSL	188	132.1 ± 323.7	179.1 ± 438.7	2,460.0 ± 5,466.5	327.6 ± 706.1

Table S8. BLASTp results of secondary metabolite genes

antiSMASH cluster	cluster gene number	antiSMASH predicted gene type	Description	Percent ID	E-value	Query cover	Accession
11	3	core biosynthetic	autoinducer synthase [Ruegeria conchae]	98.59%	0	100%	WP_010442078.1
25	25	other	16 kDa heat shock protein A [Phaeobacter italicus]	82.78%	2.00E-86	86%	CRL16738.1
32	8	core biosynthetic	autoinducer synthase [Ruegeria sp. CAU 1488]	80.34%	1.00E-171	100%	WP_138843562.1
43	2	core biosynthetic	autoinducer synthase [Ruegeria sp. AD91A]	84.62%	3.00E-147	100%	WP_117868604.1
51	63	other	hypothetical protein [Chthoniobacter flavus]	30.91%	2.00E-06	88%	WP_006978677.1
54	188	core biosynthetic	autoinducer synthase [Ruegeria sp. CAU 1488]	71.28%	5.00E-148	95%	WP_138843562.1
77	6	core biosynthetic	non-ribosomal peptide synthetase [Acidobacteria bacterium]	32.04%	0	98%	PYS22425.1
77	7	other	sigma-70 family RNA polymerase sigma factor [Coraliomargarita akajimensis]	69.36%	7.00E-84	100%	WP_013043527.1
77	8	other	DUF3520 domain-containing protein [Coraliomargarita akajimensis]	57.08%	0	98%	WP_013043526.1
82	1	other	cold-shock protein [Ruegeria sp. 6PALISEP08]	98.53%	2.00E-42	100%	WP_050602026.1
104	38	other	cold-shock protein [Vibrio tubiashii]	98.57%	1.00E-41	100%	WP_038201891.1
113	4	core biosynthetic	non-ribosomal peptide synthetase [Nostoc flagelliforme]	35.81%	0.00E+00	96%	WP_100901985.1
119	2	other	Hypothetical protein EAZ71_12040 [Verrucomicrobium bacterium]	36.65%	4.00E-35	94%	TAF23812.1
120	20	core biosynthetic	Sugar transferase [Mameliella sp.]	56.65%	1.00E-79	93%	MBT56230.1

124	5	Core biosynthetic	arthrofactin-type cyclic lipopetide synthetase C [Pseudomonas poae]	44.27%	2.00E-49	85%	SDN61670.1
125	1	core biosynthetic	amino acid adenylation domain-containing protein [Myxococcus sp. AB025A]	35.73%	1.00E-168	95%	WP_141594610.1
144	39	core biosynthetic	autoinducer synthase [Ruegeria halocynthiae]	95.17%	3.00E-144	100%	WP_074724014.1
190	6	core biosynthetic	autoinducer synthase [Ruegeria sp. ANG-S4]	99.65%	0.00E+00	100%	WP_039529872.1
192	18	other	helix-turn-helix transcriptional regulator [Pseudooceanicola atlanticus]	90.28%	9.00E-43	100%	WP_043744810.1
194	15	core biosynthetic	undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase [Clostridiales bacterium GWD2_32_19]	44.55%	4.00E-66	85%	OGO83292.1
194	16	other	TPA: sensor histidine kinase [Selenomonas sp.]	29.86%	9.00E-11	56%	HBT79610.1
194	21	core biosynthetic	Polysaccharide biosynthesis/export protein [Verrucomicrobiae bacterium DG1235]	37.21%	3.00E-33	83%	WP_008098542.1
194	24	other	PEP-CTERM sorting domain-containing protein [Geobacter sp.]	33.81%	2.00E-25	97%	RQW83478.1
200	2	core biosynthetic	AMP-binding protein [Candidatus Thiosymbion oneisti]	57.64%	1.00E-153	91%	WP_089724116.1
207	1	core biosynthetic	AMP-binding protein [Candidatus Thiosymbion oneisti]	54.01%	0.00E+00	93%	WP_089724116.1
207	4	other	VOC family protein [Verrucomicrobia bacterium]	52.54%	2.00E-33	93%	RME66880.1
211	4	core biosynthetic	FabA-like domain protein [Opitutaceae bacterium]	77.50%	5.00E-162	87%	MBE35560.1
220	22	core biosynthetic	undecaprenyl/decaprenyl-phosphate alpha-N-acetylglucosaminyl 1-	39.83%	2.00E-65	94%	WP_013374773.1

			phosphate transferase [Stigmatella aurantiaca]				
220	27	core biosynthetic	Polysaccharide biosynthesis/export protein [Verrucomicrobiae bacterium DG1235]	36.51%	6.00E-34	81%	WP_008098542.1
220	30	other	PEP-CTERM sorting domain-containing protein [Geobacter sp.]	33.09%	6.00E-23	97%	RQW83478.1
229	16	other	hypothetical protein [Cellulosilyticum ruminicola]	33.71%	0.68	58%	WP_054739542.1
253	574	other	cold-shock protein [Cribrihabitans marinus]	92.54%	1.00E-39	98%	WP_092362673.1

Table S9. Significantly different average gene expression comparisons for secondary metabolite biosynthetic genes. Adjusted *p* values are from a mixed-effects model analysis with Tukey multiple comparisons. All other comparisons from heatmaps in Fig. 2-4 are not significantly different. Top blastp hit for each gene is also shown.

Gene information				Multiple comparisons testing			Top blastp result			
Heat map	**Heat map row**	**Cluster number**	**Cluster gene number**	**Comparison**	**Mean Diff.**	**Adjusted *p* value**	**Function**	**Organism**	**%ID**	**E-value**
HSL	22	171	279	ANG vs. Day 10 JC Control	-278.1	0.0207	SDR family NAD(P)-dependent oxidoreductase	Ruegeria sp. ANG-R	98.41	1.00E-175
NRPS	1	220	22	Day 0 JC vs. Day 10 JC Control	90711	0.0198	undecaprenyl/decaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase	Stigmatella aurantiaca	39.83	2.00E-65
NRPS	1	220	22	Day 0 JC vs. Day 10 JC Challenge	85889	0.0109	undecaprenyl/decaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase	Stigmatella aurantiaca	39.83	2.00E-65
NRPS	2	194	15	Day 0 JC vs. Day 10 JC Control	42712	0.0495	undecaprenyl/decaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase	Clostridiales bacterium GWD2_32_19	44.55	4.00E-66
NRPS	2	194	15	Day 0 JC vs. Day 10 JC Challenge	42713	0.0295	undecaprenyl/decaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase	Clostridiales bacterium GWD2_32_19	44.55	4.00E-66

Table S10. Top 20 most highly expressed genes from the *Verrucomicrobia* PD13 Bin 1 MAG and those genes which are high amongst the ANG and JC metatranscriptomes

Samples	# of shared genes	RAST Gene ID	Function	Top blastx hit	Percent ID	E-value	Accession
ANG, Day 0 JC , Day 10 JC Control, Day 10 JC Challenge	11	6666666.429098.peg.2230	MFS transporter	MFS transporter [Puniceicoccaceae bacterium]	43.7	3.40E-78	RCL36448.1
		6666666.429098.peg.1585	UPF0301 protein YqgE	YqgE/AlgH family protein [Verrucomicrobia bacterium]	53.7	2.40E-47	RME73046.1
		6666666.429098.peg.1240	Acyltransferase	acyltransferase [Gramella flava]	57.1	1.00E-38	WP_083644399.1
		6666666.429098.peg.3214	Glycosyl hydrolase	Family 43 glycosylhydrolase [Echinicola pacifica]	66	9.7E-147	WP_018475183.1
		6666666.429098.peg.1698	DUF3303 domain-containing protein	DUF3303 domain-containing protein [Cupriavidus necator]	39.8	1.6E-09	WP_114130888.1
		6666666.429098.peg.3682	Hypothetical protein	ABC transporter C family member 8 [Ananas comosus]	27	8	OAY83979.1
		6666666.429098.peg.1070	PBS lyase HEAT domain-containing protein	PBS lyase HEAT domain-containing protein [Coraliomargarita akajimensis]	60.5	0	WP_013042711.1
		6666666.429098.peg.745	mandelate racemase/muconate lactonizing enzyme family protein	Mandelate racemase/muconate lactonizing enzyme family protein [Porphyrobacter sp. HT-58-2]	57	2.9E-136	WP_104941388.1
		6666666.429098.peg.755	sugar phosphate isomerase/epimerase	sugar phosphate isomerase/epimerase [Lentisphaera araneosa]	56	8.5E-79	WP_007279462.1
		6666666.429098.peg.2790	Thiol peroxidase, Bcp-type	peroxiredoxin [Prolixibacteraceae bacterium XSD2]	56.6	1.6E-49	WP_119437511.1

		6666666.429098.peg.829	GH81, signal peptide	carbohydrate-binding protein [Polaribacter gangjinensis]	57.9	0	WP_105045425.1
ANG, Day 0 JC, Day 10 JC Control	1	6666666.429098.peg.3171	Putative stomatin/prohibitin-family membrane protease subunit PA4582	peptidase [Candidatus Lindowbacteria bacterium]	50	1.5E-88	OGH56740.1
ANG, Day 10 JC Control, Day 10 JC Challenge	2	6666666.429098.peg.1876	Phage shock protein A, Suppressor of sigma54-dependent transcription, PspA-like	phage shock protein A [Nitrospira bacterium SG8_3]	64.8	1.9E-62	KPK29165.1
		6666666.429098.peg.3441	Membrane protease family protein BA0301	band 7 protein [Verrucomicrobia bacterium]	64.8	1.5E-111	PYJ02783.1
Day 0 JC, Day 10 JC Control, Day 10 JC Challenge	2	6666666.429098.peg.593	Response regulator	response regulator [Coraliomargarita akajimensis]	47.1	1.1E-47	WP_013043653.1
		6666666.429098.peg.1414	Hypothetical protein	hypothetical protein DME86_05660 [Verrucomicrobia bacterium]	50.9	5.6E-22	PYJ42333.1
ANG, Day 0 JC	3	6666666.429098.peg.1843	Nitrogen regulatory protein P-II	P-II family nitrogen regulator [Opitutaceae bacterium TAV5]	77.7	1.7E-43	WP_009513939.1
		6666666.429098.peg.457	SOS-response repressor and protease LexA (EC 3.4.21.88)	LexA repressor [Candidatus Moanabacter tarae]	75.4	2.5E-88	AWT60219.1
		6666666.429098.peg.1274	DNA-binding protein HU-beta	DNA-binding protein [Opitutaceae bacterium BACL24 MAG-120322-bin51]	79.5	2.5E-29	KRP36679.1
Day 10 JC Control, Day 10 JC Challenge	4	6666666.429098.peg.1632	RNA polymerase sigma factor RpoE	sigma-70 family RNA polymerase sigma factor [Coraliomargarita sp. TMED73]	54.7	1.3E-47	RPG85051.1

		6666666.429098.peg.4028	Hypothetical protein	hypothetical protein [Vibrio owensii]	44.3	8.9E-07	WP_052470655.1
		6666666.429098.peg.476	Hypothetical protein	hypothetical protein DF168_01663 [Candidatus Moanabacter tarae]	33.2	4.3E-125	AWT60449.1
		6666666.429098.peg.4033	Hypothetical protein	no significant match	NA	NA	NA
ANG	3	6666666.429098.peg.1184	Hypothetical protein	hypothetical protein A3G75_14710 [Verrucomicrobia bacterium]	49.5	3.4E-17	OHE83810.1
		6666666.429098.peg.2837	Acyl carrier protein	acyl carrier protein [Verrucomicrobia bacterium]	81.2	2.2E-27	OHE79247.1
		6666666.429098.peg.1756	Peptide transporter	peptide transporter [Verrucomicrobia bacterium LW23]	60.4	2.7E-226	WP_108035984.1
Day 0 JC	3	6666666.429098.peg.3091	Hypothetical protein	hypothetical protein A2107_05890 [Verrucomicrobia bacterium GWF2_62_7]	55.2	5E-72	OHE75575.1
		6666666.429098.peg.258	GNAT family N-acetyltransferase	GNAT family N-acetyltransferase [Indioceanicola profundi]	29.6	6.5E-06	WP_119679184.1
		6666666.429098.peg.341	Hypothetical protein	hypothetical protein DRP71_08335 [Verrucomicrobia bacterium]	54.9	2.3E-71	RKX34075.1
Day 10 JC Challenge	1	6666666.429098.peg.937	Hypothetical protein	GAF domain-containing sensor histidine kinase [Gammaproteobacteria bacterium]	50	7.9	TLZ04684.1

Table S11. Significantly different average gene expression comparisons for *Verrucomicrobia* PD13 Bin 1 genes. Adjusted *p* values are from a mixed-effects model analysis with Tukey multiple comparisons. All other comparisons from heatmaps in Fig. 7-8 are not significantly different.

Heat map	Heat map row	Gene ID	function	Comparison	Mean Diff.	Adjusted P Value
Fucose	15	6666666.429098.peg.2081	alpha-L-fucosidase	ANG vs. Day 10 JC Control	-237.8	0.021
Fucose	24	6666666.429098.peg.3952	alpha-L-fucosidase	ANG vs. Day 0 JC	42.1	0.042
Fucose	24	6666666.429098.peg.3952	alpha-L-fucosidase	ANG vs. Day 10 JC Control	56.94	0.0166
Fucose	27	6666666.429098.peg.544	L-fucose isomerase	ANG vs. Day 0 JC	179.7	0.0486
Fucose	27	6666666.429098.peg.544	L-fucose isomerase	ANG vs. Day 10 JC Control	316.4	0.032
Fucose	27	6666666.429098.peg.544	L-fucose isomerase	ANG vs. Day 10 JC Challenge	286.8	0.0432
Fucose	28	6666666.429098.peg.667	L-fucose kinase	Day 10 JC Control vs. Day 10 JC Challenge	-5.816	0.0061
Fucose	30	6666666.429098.peg.2583	GDP-L-fucose synthetase	Day 0 JC vs. Day 10 JC Control	-7.821	0.0457
Sialic acid	2	6666666.429098.peg.724	Predicted sialic acid transporter	Day 0 JC vs. Day 10 JC Challenge	-4.215	0.0354
Sulfur utilization	5	6666666.429098.peg.3046	Cysteine synthase	ANG vs. Day 0 JC	-2324	0.0289
Sulfur utilization	7	6666666.429098.peg.1908	Ferredoxin	Day 10 JC control vs. Day 10 JC challenge	-22.95	0.0189
Sulfur utilization	11	6666666.429098.peg.322	Iron-sulfur cluster-binding protein	Day 0 JC vs. Day 10 JC challenge	-63.82	0.0241
Sulfur utilization	16	6666666.429098.peg.3821	Phosphoadenylyl-sulfate reductase	ANG vs. Day 0 JC	-241.3	0.0456
Sulfur utilization	16	6666666.429098.peg.3821	Phosphoadenylyl-sulfate reductase	ANG vs. Day 10 JC control	-376.1	0.0291
Sulfur utilization	26	6666666.429098.peg.283	Sulfide:quinone oxidoreductase, Type III	ANG vs. Day 10 JC control	309.5	0.0475

Sulfur utilization	26	6666666.429098.peg.1155	Sulfur carrier protein FdhD	Day 0 JC vs. Day 10 JC challenge	-11.54	0.0152
Motility & chemotaxis	10	6666666.429098.peg.1468	Flagellar basal-body rod protein FlgG	ANG vs. Day 10 JC Control	12.83	0.0157
Motility & chemotaxis	22	6666666.429098.peg.1524	Flagellum biosynthesis repressor protein FlbT	ANG vs. Day 10 JC Challenge	10.36	0.0084
TIV Pili	1	6666666.429098.peg.2419	Type IV fimbrial assembly protein PilC	Day 0 JC vs. Day 10 JC Control	143.8	0.0417
TIV Pili	1	6666666.429098.peg.2419	Type IV fimbrial assembly protein PilC	Day 0 JC vs. Day 10 JC Challenge	111.1	0.0161
TIV Pili	2	6666666.429098.peg.1575	Type IV fimbrial assembly, ATPase PilB	Day 0 JC vs. Day 10 JC Challenge	-69.07	0.0241
TIV Pili	5	6666666.429098.peg.3806	Type IV fimbrial assembly, ATPase PilB	Day 0 JC vs. Day 10 JC Challenge	-34.85	0.0382
Transport	29	6666666.429098.peg.1857	ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component	ANG vs. Day 0 JC	7.698	0.0273
Transport	31	6666666.429098.peg.1168	Acidobacterial duplicated orphan permease	Day 0 JC vs. Day 10 JC Control	-71.66	0.0215
Transport	31	6666666.429098.peg.1168	Acidobacterial duplicated orphan permease	Day 0 JC vs. Day 10 JC Challenge	-95.81	0.0206
Transport	33	6666666.429098.peg.3974	Acidobacterial duplicated orphan permease	ANG vs. Day 0 JC	34.81	0.0432
Transport	34	6666666.429098.peg.3975	Acidobacterial duplicated orphan permease	ANG vs. Day 0 JC	35.04	0.0459
Transport	37	6666666.429098.peg.887	AttH of AttEFGH ABC transport system	Day 0 JC vs. Day 10 JC Control	-11.9	0.0379
Transport	60	6666666.429098.peg.3006	Ferrous iron transport protein B	Day 0 JC vs. Day 10 JC Challenge	-58.49	0.0353
Transport	71	6666666.429098.peg.632	L-rhamnose-proton symporter	Day 0 JC vs. Day 10 JC Control	51.61	0.0043
Transport	77	6666666.429098.peg.2043	MotA/TolQ/ExbB proton channel family protein	ANG vs. Day 10 JC Control	55.03	0.0123
Transport	87	6666666.429098.peg.1673	Na(+) H(+) antiporter subunit D	Day 0 JC vs. Day 10 JC Challenge	-10.31	0.0207

Transport	93	6666666.429098.peg.2272	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	Day 0 JC vs. Day 10 JC Challenge	-54.22	0.0198
Transport	97	6666666.429098.peg.2269	Oligopeptide transport ATP-binding protein OppD	Day 10 JC Control vs. Day 10 JC Challenge	-11.19	0.0106
Transport	118	6666666.429098.peg.724	Predicted sialic acid transporter	Day 0 JC vs. Day 10 JC Challenge	-4.215	0.0354
Transport	125	6666666.429098.peg.610	Putative nucleoside transporter yegT	ANG vs. Day 10 JC Challenge	24.16	0.0296
Transport	127	6666666.429098.peg.3908	Putative transport protein	Day 0 JC vs. Day 10 JC Control	-20.72	0.0101
Transport	131	6666666.429098.peg.846	RND efflux system, inner membrane transporter	ANG vs. Day 0 JC	17.03	0.016
Transport	131	6666666.429098.peg.846	RND efflux system, inner membrane transporter	Day 0 JC vs. Day 10 JC Challenge	-43.16	0.011
Transport	137	6666666.429098.peg.2274	Sodium-dependent transporter, SNF family	Day 0 JC vs. Day 10 JC Control	-20.3	0.0447
Transport	145	6666666.429098.peg.2431	TonB-dependent receptor	Day 0 JC vs. Day 10 JC Challenge	-6.138	0.0183
Transport	147	6666666.429098.peg.7	TonB-dependent receptor plug	ANG vs. Day 10 JC Control	663.7	0.0346
Transport	148	6666666.429098.peg.1139	Transporter, sodium/sulfate symporter family	Day 10 JC Control vs. Day 10 JC Challenge	-6.85	0.0442
Transport	149	6666666.429098.peg.1150	TRAP-type C4-dicarboxylate transport system, large permease component	Day 0 JC vs. Day 10 JC Challenge	-9.888	0.0282
Transport	152	6666666.429098.peg.3861	Trk potassium uptake system protein TrkH	Day 0 JC vs. Day 10 JC Challenge	-16.53	0.0299
Transport	157	6666666.429098.peg.1031	Uncharacterized MFS-type transporter	Day 0 JC vs. Day 10 JC Challenge	-12.81	0.0011
Sulfatase	6	6666666.429098.peg.644	Arylsulfatase	ANG vs. Day 10 JC Challenge	11.13	0.0298
Sulfatase	7	6666666.429098.peg.707	Arylsulfatase	Day 0 JC vs. Day 10 JC Control	-20.08	0.0439
Sulfatase	7	6666666.429098.peg.707	Arylsulfatase	Day 0 JC vs. Day 10 JC Challenge	-21.13	0.0209

Sulfatase	8	6666666.429098.peg.2708	Arylsulfatase	Day 0 JC vs. Day 10 JC Challenge	-24.22	0.008
Sulfatase	32	6666666.429098.peg.227	Arylsulfatase	Day 0 JC vs. Day 10 JC Challenge	17.42	0.0036
Sulfatase	65	6666666.429098.peg.715	Arylsulfatase	Day 10 JC Control vs. Day 10 JC Challenge	-6.39	0.0033
Sulfatase	115	6666666.429098.peg.1607	Arylsulfatase	Day 0 JC vs. Day 10 JC Challenge	-30.41	0.038
Sulfatase	115	6666666.429098.peg.1607	Arylsulfatase	Day 10 JC Control vs. Day 10 JC Challenge	-7.456	0.0086
Sulfatase	132	6666666.429098.peg.1818	Arylsulfatase	Day 0 JC vs. Day 10 JC Challenge	-6.875	0.0015
Sulfatase	136	6666666.429098.peg.2084	Arylsulfatase	ANG vs. Day 10 JC Challenge	-10.99	0.0379
Sulfatase	142	6666666.429098.peg.2705	Arylsulfatase	Day 0 JC vs. Day 10 JC Challenge	-28.91	0.0009
Sulfatase	145	6666666.429098.peg.2743	Arylsulfatase	Day 0 JC vs. Day 10 JC Control	-9.62	0.0407
Sulfatase	157	6666666.429098.peg.3192	Arylsulfatase	Day 0 JC vs. Day 10 JC Challenge	-6.285	0.0069
Sulfatase	170	6666666.429098.peg.3330	Arylsulfatase	Day 0 JC vs. Day 10 JC Control	-9.243	0.0271
Sulfatase	171	6666666.429098.peg.3331	Arylsulfatase	Day 0 JC vs. Day 10 JC Control	-9.057	0.0129
Sulfatase	172	6666666.429098.peg.3340	Arylsulfatase	Day 0 JC vs. Day 10 JC Control	-5.545	0.0092
Sulfatase	172	6666666.429098.peg.3340	Arylsulfatase	Day 0 JC vs. Day 10 JC Challenge	-7.655	0.0226
Sulfatase	176	6666666.429098.peg.3374	Arylsulfatase	ANG vs. Day 0 JC	4.55	0.0256
Sulfatase	176	6666666.429098.peg.3374	Arylsulfatase	ANG vs. Day 10 JC Control	4.045	0.0451
Sulfatase	193	6666666.429098.peg.1100	Putative protein related to sulfatases	Day 0 JC vs. Day 10 JC Challenge	-10.86	0.0457
Sulfatase	211	6666666.429098.peg.641	Sulfatase	Day 0 JC vs. Day 10 JC Challenge	-4.82	0.009

Sulfatase	212	6666666.429098.peg.647	Sulfatase	Day 0 JC vs. Day 10 JC Challenge	-6.441	0.0341
Sulfatase	218	6666666.429098.peg.851	Sulfatase	Day 0 JC vs. Day 10 JC Control	-5.507	0.0241
Sulfatase	245	6666666.429098.peg.2788	Sulfatase	Day 0 JC vs. Day 10 JC Control	-14.02	0.0248
Sulfatase	245	6666666.429098.peg.2788	Sulfatase	Day 0 JC vs. Day 10 JC Challenge	-14.9	0.0313
Sulfatase	247	6666666.429098.peg.2928	Sulfatase	Day 0 JC vs. Day 10 JC Control	19.65	0.0026
Sulfatase	158	6666666.429098.peg.3414	Sulfatase	Day 0 JC vs. Day 10 JC Challenge	-7.162	0.0102
Carbohydrate active enzymes	56	6666666.429098.peg.2691	CE16 / hypothetical protein	ANG vs. Day 10 JC Challenge	80.33	0.0279
Carbohydrate active enzymes	69	6666666.429098.peg.3222	GH82 / hypothetical protein	Day 0 JC vs. Day 10 JC Control	-8.5	0.0175
Carbohydrate active enzymes	75	6666666.429098.peg.1715	GH10 / hypothetical protein	Day 0 JC vs. Day 10 JC Challenge	-13.93	0.0094
Carbohydrate active enzymes	89	6666666.429098.peg.918	GH109 / probable NADH-dependent dyhydrogenase	Day 0 JC vs. Day 10 JC Control	-6.102	0.0281
Carbohydrate active enzymes	107	6666666.429098.peg.3221	GH109 / hypothetical protein	Day 0 JC vs. Day 10 JC Challenge	56.75	0.0043
Carbohydrate active enzymes	118	6666666.429098.peg.3251	GH109 / hypothetical protein	Day 10 JC Control vs. Day 10 JC Challenge	-13.18	0.014
Carbohydrate active enzymes	135	6666666.429098.peg.862	GH117 / hypothetical protein	ANG vs. Day 0 JC	52.12	0.013
Carbohydrate active enzymes	137	6666666.429098.peg.1047	GH117 / hypothetical protein	ANG vs. Day 10 JC Control	341.1	0.0411

Carbohydrate active enzymes	137	6666666.429098.peg.1047	GH117 / hypothetical protein	ANG vs. Day 10 JC Challenge	292.2	0.0404
Carbohydrate active enzymes	141	6666666.429098.peg.2955	GH117 / hypothetical protein	Day 0 JC vs. Day 10 JC Control	369.4	0.0361
Carbohydrate active enzymes	143	6666666.429098.peg.3226	GH117 / hypothetical protein	Day 0 JC vs. Day 10 JC Control	153.4	0.0007
Carbohydrate active enzymes	143	6666666.429098.peg.3226	GH117 / hypothetical protein	Day 0 JC vs. Day 10 JC Challenge	146.8	0.0242
Carbohydrate active enzymes	150	6666666.429098.peg.853	GH127 / Neopullulanase (EC 3.2.1.135)	Day 0 JC vs. Day 10 JC Control	-7.913	0.0352
Carbohydrate active enzymes	150	6666666.429098.peg.853	GH127 / Uncharacterized glycosyltransferase YcjM	Day 0 JC vs. Day 10 JC Challenge	-8.817	0.0243
Carbohydrate active enzymes	161	6666666.429098.peg.1166	GH13 / Arylsulfatase (EC 3.1.6.1)	ANG vs. Day 10 JC Challenge	-51.19	0.0469
Carbohydrate active enzymes	163	6666666.429098.peg.1582	GH13 / Arylsulfatase (EC 3.1.6.1)	ANG vs. Day 10 JC Challenge	-247.6	0.0276
Carbohydrate active enzymes	197	6666666.429098.peg.644	GH16 / Arylsulfatase (EC 3.1.6.1)	ANG vs. Day 10 JC Challenge	11.13	0.0298
Carbohydrate active enzymes	198	6666666.429098.peg.707	GH16 / Arylsulfatase (EC 3.1.6.1)	Day 0 JC vs. Day 10 JC Control	-20.08	0.0439
Carbohydrate active enzymes	198	6666666.429098.peg.707	GH16 / Beta-glucanase precursor (EC 3.2.1.73)	Day 0 JC vs. Day 10 JC Challenge	-21.13	0.0209
Carbohydrate active enzymes	199	6666666.429098.peg.2708	GH16 / Beta-glucanase precursor (EC 3.2.1.73)	Day 0 JC vs. Day 10 JC Challenge	-24.22	0.008

Carbohydrate active enzymes	200	6666666.429098.peg.1898	GH16 / Beta-glucanase precursor (EC 3.2.1.73)	Day 0 JC vs. Day 10 JC Control	-29.67	0.0093
Carbohydrate active enzymes	201	6666666.429098.peg.1913	GH16 / beta-galactosidase (EC 3.2.1.23)	ANG vs. Day 10 JC Control	-188.1	0.0489
Carbohydrate active enzymes	201	6666666.429098.peg.1914	GH17 / alpha-L-fucosidase (EC 3.2.1.51)	ANG vs. Day 10 JC Challenge	-279.6	0.0106
Carbohydrate active enzymes	219	6666666.429098.peg.976	GH2 / alpha-L-fucosidase (EC 3.2.1.51)	ANG vs. Day 10 JC Challenge	12.4	0.0489
Carbohydrate active enzymes	234	6666666.429098.peg.625	GH2 / alpha-L-fucosidase (EC 3.2.1.51)	Day 10 JC Control vs. Day 10 JC Challenge	-3.43	0.02
Carbohydrate active enzymes	261	6666666.429098.peg.2081	GH29 / alpha-L-fucosidase (EC 3.2.1.51)	ANG vs. Day 10 JC Control	-237.8	0.021
Carbohydrate active enzymes	269	6666666.429098.peg.3951	GH29 / hypothetical protein	ANG vs. Day 0 JC	42.1	0.042
Carbohydrate active enzymes	269	6666666.429098.peg.3951	GH29 / hypothetical protein	ANG vs. Day 10 JC Control	56.94	0.0166
Carbohydrate active enzymes	272	6666666.429098.peg.2208	GH3 / alpha-galactosidase (EC 3.2.1.22)	Day 0 JC vs. Day 10 JC Challenge	-7.935	0.0144
Carbohydrate active enzymes	281	6666666.429098.peg.1652	GH31 / alpha-galactosidase (EC 3.2.1.22)	Day 10 JC Control vs. Day 10 JC Challenge	-5.534	0.0208
Carbohydrate active enzymes	305	6666666.429098.peg.2733	GH4 / hypothetical protein	Day 0 JC vs. Day 10 JC Control	-53.43	0.0235
Carbohydrate active enzymes	305	6666666.429098.peg.2733	GH4 / putative cellulase family	Day 0 JC vs. Day 10 JC Challenge	-74.67	0.0319

Carbohydrate active enzymes	317	6666666.429098.peg.1702	GH49 / hypothetical protein	Day 10 JC Control vs. Day 10 JC Challenge	10.12	0.0168
Carbohydrate active enzymes	322	6666666.429098.peg.3434	GH5 / hypothetical protein	Day 0 JC vs. Day 10 JC Control	-73.64	0.0335
Carbohydrate active enzymes	335	6666666.429098.peg.649	GH51 / hypothetical protein	Day 0 JC vs. Day 10 JC Control	-314.5	0.0046
Carbohydrate active enzymes	335	6666666.429098.peg.649	GH51 / 4-alpha-glucanotransferase (amylomaltase) (EC 2.4.1.25)	Day 0 JC vs. Day 10 JC Challenge	-298	0.0069
Carbohydrate active enzymes	336	6666666.429098.peg.3088	GH51 / Putative large secreted protein SCO0341	Day 0 JC vs. Day 10 JC Challenge	-233.8	0.0377
Carbohydrate active enzymes	343	6666666.429098.peg.1164	GH77 / hypothetical protein	Day 0 JC vs. Day 10 JC Challenge	-35.95	0.0354
Carbohydrate active enzymes	366	6666666.429098.peg.849	GH95 / Bactoprenol glucosyl transferase	Day 0 JC vs. Day 10 JC Challenge	-7.971	0.0414
Carbohydrate active enzymes	370	6666666.429098.peg.3227	GH97 / Bactoprenol glucosyl transferase	Day 0 JC vs. Day 10 JC Challenge	-21.82	0.0313
Carbohydrate active enzymes	374	6666666.429098.peg.2500	GT2 / Polyketide synthase modules and related proteins	Day 0 JC vs. Day 10 JC Control	122.5	0.026
Carbohydrate active enzymes	374	6666666.429098.peg.2500	GT2 / hypothetical protein	Day 0 JC vs. Day 10 JC Challenge	115.1	0.0271
Carbohydrate active enzymes	387	6666666.429098.peg.1173	GT2 / hypothetical protein	Day 10 JC Control vs. Day 10 JC Challenge	-3.368	0.0485
Carbohydrate active enzymes	390	6666666.429098.peg.1346	GT26 / UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide)	Day 0 JC vs. Day 10 JC Control	132.2	0.0002

			pyrophosphoryl-undecaprenol N-acetylglucosamine transferase (EC 2.4.1.227)			
Carbohydrate active enzymes	390	6666666.429098.peg.1346	GT26 / hypothetical protein	Day 0 JC vs. Day 10 JC Challenge	99.98	0.0023
Carbohydrate active enzymes	392	6666666.429098.peg.3134	GT28 / hypothetical protein	Day 0 JC vs. Day 10 JC Control	-34.83	0.0483

Table S12. Significantly differentially expressed genes in *Verrucomicrobia* PD13 Bin 1 MAG mapped reads. These values correspond to red points on the volcano plots in Figures 9 and S3. Blastp results for hypothetical proteins shown.

Comparison	Sample 1 mean	Sample 2 mean	theta	log2 FC	P-value	RAST annotation	subject name	Percent ID	E-value	subject ID
ANG v. Day 0 JCs	565.71	2715.21	-2.26	-2.26	0.034	Hypothetical protein	Cell wall hydrolase [Coraliomargarita akajimensis]	62.1	3.00E-38	WP_013044581.1
	35.94	1496.70	-2.18	-5.37	0.046	Hypothetical protein	Fasciclin domain-containing protein [Aquimarina amphilecti]	41.5	1.20E-35	WP_091411286.1
Day 0 JCs v. Day 10 control JCs	10575.26	4811.17	2.73	1.13	1.47E-14	Hypothetical protein	NA	NA	NA	NA
	10547.94	4146.08	2.56	1.34	5.97E-07	Undecaprenyl-phosphate alpha-N-acetyl-glucosaminyl 1-phosphate transferase	NA	NA	NA	NA
	3991.62	1728.33	2.49	1.20	0.00092	Hypothetical protein	GHKL domain-containing protein [Blautia hansenii]	25.9	1.50E-08	WP_009246618.1
	2470.59	301.72	2.49	3.03	0.0012	Hypothetical protein	Hypothetical protein B9S34_11585 [Opitutae bacterium Tous-C1TDCM]	36.5	2.30E-19	PAW65034.1
Day 10 control JCs v. Day 10 challenge JCs	5.22	10.79	-0.30	-1.04	0.049	hypothetical protein	NA	NA	NA	NA
	4.32	1.89	0.28	1.18	0.049	hypothetical protein	NA	NA	NA	NA
	1.49	0.73	0.27	1.02	0.046	hypothetical protein	NA	NA	NA	NA

ANG v. Day 10 control JCs	1030.41	41.85	2.85	4.62	1.35E-09	hypothetical protein	NA	NA	NA	NA
	276.22	934.69	-1.82069	-1.75	0.032	ABC-type antimicrobial peptide transport	NA	NA	NA	NA
	3268.55	9553.59	-1.83687	-1.54	0.033	hypothetical protein	NA	NA	NA	NA
	717.20	1998.87	-2.20654	-1.47	0.033	8-amino-7-oxononanoate synthase				
	1557.83	2983.67	-1.80236	-0.93	0.035	Thioredoxin reductase				
	2580.79	594.96	2.572381	2.11	0.037	TonB-dependent receptor	NA	NA	NA	NA
	1095.32	6566.86	-1.85339	-2.58	0.038	hypothetical protein	NA	NA	NA	NA
	3788.72	144.48	2.578175	4.71	0.041	hypothetical protein	NA	NA	NA	NA
	1108.23	3045.51	-1.7706	-1.45	0.043	Protein yceI precursor	NA	NA	NA	NA
	205.98	731.67	-1.76097	-1.82	0.045	hypothetical protein	NA	NA	NA	NA
Day 0 JCs v. Day 10 challenge JCs	3960.14	1684.17	2.581636	1.23	2.71E-10	hypothetical protein	NA	NA	NA	NA
	2451.11	288.05	2.545744	3.08	8.90E-08	hypothetical protein	NA	NA	NA	NA
	3309.89	865.65	2.099443	1.93	0.022	hypothetical protein	NA	NA	NA	NA
	536.02	49.01	2.094391	3.45	0.023	hypothetical protein	NA	NA	NA	NA
	10464.75	4573.82	2.247588	1.19	0.024	Undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase	NA	NA	NA	NA
	2678.58	693.11	2.062502	1.95	0.029	hypothetical protein	NA	NA	NA	NA
	6363.22	947.53	2.05369	2.74	0.031	DNA-binding protein HU-beta	NA	NA	NA	NA
	24145.04	7911.55	1.995021	1.61	0.046	hypothetical protein	NA	NA	NA	NA
	1415.11	278.76	1.812158	2.34	0.046	hypothetical protein	NA	NA	NA	NA
	1390.65	323.01	1.809701	2.106101	0.047	hypothetical protein	NA	NA	NA	NA
ANG v. Day 10 challenge JCs	5977.77	288.1883	4.024468	4.374526	6.38E-14	hypothetical protein	NA	NA	NA	NA
	3749.52	121.92	2.75	4.94	0.00049	hypothetical protein	NA	NA	NA	NA
	1019.75	48.94	2.59	4.38	0.04	hypothetical protein	NA	NA	NA	NA

Chapter 5

Conclusions and Future Directions

The bacterial symbionts of the ANG play a role in egg defense over the course of embryo development in *Euprymna scolopes*. While it has been shown that a metabolically active jelly coat community is necessary to protect against fungal/bacterial infections in eggs (Appendix I), the mechanisms behind this protection have not been described. This dissertation has demonstrated that ANG/JC symbionts can inhibit marine bacteria and fungi *in vitro* and express biosynthetic genes for several classes of secondary metabolites, including nonribosomal peptide synthetases, polyketide synthases, and bacteriocins. These classes of metabolites are known to produce defensive bioactive compounds in other symbioses (31), but the exact compounds produced in the ANG/eggs remains to be characterized, although some known and putative compounds have been identified (85)(Chapter 1, Appendix I) . With the symbionts already in culture, biosynthetic gene clusters that were similar to those expressed in the ANG/JC can be knocked out to assess if antimicrobial activity is also lost. Mutants lacking these secondary metabolite gene clusters could be used to colonize ANGs and determine if eggs that receive these symbionts are more vulnerable to infections.

The similar expression of secondary metabolite biosynthetic genes in the ANG and egg JCs suggests that bioactive compounds are continuously made by the bacterial symbionts. Although the exact compounds are not known, short half-lives or continuous diffusion into the surrounding seawater may necessitate that these compounds be produced continuously while the bacteria are in association with the host. Comparative studies of gene expression under host-like conditions versus free-living, seawater conditions in symbionts may reveal that some host factors are required for heightened expression of secondary metabolites. To gain a better understanding of what is regulating bacterial gene expression, however, more research is needed to understand the host microenvironments of both the ANG and eggs.

Analyzing the ANG metagenomes and ANG/JC metatranscriptomes revealed new information about symbionts that have not yet been cultured. Metagenome-assembled genomes (MAGs) were recovered from several *Verrucomicrobia* strains, as well as *Alterythrobacter*, *Mesorhizobium*, and *Vibrio* strains. While only one *Verrucomicrobia* strain was discussed in depth in this dissertation, examining gene expression of this MAG revealed some of the carbon sources that this strain may be utilizing in the ANG and eggs. As has been done previously (232), using this metatranscriptome data to guide media development may allow for the culturing of *Verrucomicrobia* and other currently uncultured ANG/JC community members.

A well-timed and well-researched chemical dialogue occurs between *Vibrio fischeri* and the light organ tissue in *E. scolopes*, and parallels from this symbiosis may also occur in the ANG symbiosis. Studies of the pH, viscosity, and immune molecules in the light organ tissue have all revealed important factors that drive colonization events for *V. fischeri* and the conditions under which it produces light [reviewed in (89)(94)(233)]. Characterizing the micro-environment of the ANG tubules and egg jelly coat will be important next steps in complimenting the bacterial metatranscriptomes. While this dissertation has focused primarily on prokaryotic gene expression, host transcripts were also generated for matching eukaryotic ANG samples. More analyses will be needed to examine these host transcripts for biosynthesis genes of other potential symbiont nutrient sources. These transcripts could also be used to better characterize mucin production in the ANG.

This research has provided pathways, both in symbiont primary and secondary metabolism, that can serve as the focus for future research into host colonization and community function. Metagenomic sequencing provided new genomes which allowed for a more complete

picture of the community's metabolic potential, while coupled metatranscriptomic sequencing revealed which pathways are most highly expressed in the ANG.

Appendix I

Shielding the next generation: symbiotic bacteria from a reproductive organ use chemical defense to protect eggs from fungal fouling

Significant contributions:

Significant contributions to this chapter were made by other researchers. Allison Kerwin (Department of Molecular and Cell Biology, University of Connecticut) performed antibiotic clutch experiments, egg component experiments, TEM, 16S rRNA gene sequencing and wrote portions of the paper. Sal Frasca (Department of Pathobiology and Veterinary Science, University of Connecticut) and Deanna Sutton (Fungal Testing Laboratory, San Antonio, TX), isolated and identified *F. keratoplasticum*, and provided an additional laboratory strain for extract testing. Kerry O'Donnell (National Center for Agricultural Utilization Research, Agricultural Research Service, Peoria, IL), completed MLST of the *F. keratoplasticum* isolates from CT and HI, provided the fungal phylogeny, and wrote a portion of the paper. Samantha Gromek and Marcy Balunas (Department of Pharmaceutical Sciences, University of Connecticut) obtained organic extracts, performed extract antifungal assays, generated molecular networks, and wrote portions of the paper. Spencer Nyholm (Department of Molecular and Cell Biology, University of Connecticut) contributed to experimental design, performed clutch experiments in HI, and wrote portions of the paper. My contribution was the isolation of ANG/JC strains used for organic extractions, performing antifungal well-diffusion assays, repeating antibiotic clutch experiments, obtaining egg samples for mass spec, and writing portions of the paper.

Abstract

The importance of defensive symbioses, whereby microbes protect hosts through the production of specific compounds, is becoming increasingly more evident. Although defining the partners in these associations has become easier, assigning function to these relationships often presents a significant challenge. Here we describe a functional role for a bacterial consortium in a female reproductive organ in the Hawaiian bobtail squid, *Euprymna scolopes*. Bacteria from the accessory nidamental gland (ANG) are deposited into the egg jelly coat (JC) where they are hypothesized to play a defensive role during embryogenesis. Eggs treated with an antibiotic cocktail developed a biomass primarily composed of the pathogenic fungus *Fusarium keratoplasticum* that infiltrated the JC, resulting in severely reduced hatch rates. Experimental manipulation of the eggs demonstrated that the JC was protective against this fungal fouling. A large proportion of the bacterial strains isolated from the ANG or JC inhibited *F. keratoplasticum* in culture (87.5%), while a similar proportion of extracts from these strains also exhibited antifungal activity against *F. keratoplasticum* and/or the human pathogenic yeast *Candida albicans* (72.7%). Molecular network analyses of active extracts of bacterial isolates and egg clutches revealed compounds likely involved with preventing microbial overgrowth. Several secondary metabolites were identified from ANG/JC bacteria and egg clutches, including lincomycins, glycerophosphocholines, and mycinamicin-like compounds, some of which inhibited *F. keratoplasticum in vitro*. These results shed light on a widely distributed but poorly understood symbiosis in cephalopods and offer a new source to explore bacterial secondary metabolites with antimicrobial activity.

Introduction

Defensive symbioses are found in a number of host-microbe associations, wherein secondary metabolites derived from beneficial microbes are often used to inhibit other microorganisms or to protect the host from predation (54, 234–237). This phenomenon has been well studied in insect associations, for example in a number of beetle species, termites, fungus-farming ants and in the protection of beewolf larvae (6, 31). Many organisms lay their eggs in an environment where successful embryogenesis depends on minimizing fouling by microorganisms. Aquatic organisms are especially susceptible to fouling because their eggs are under constant exposure to high densities of microorganisms, including over 10^6 bacterial cells/mL and 10^3 fungal spores/mL in most coastal seawater (238). Given that cephalopod embryogenesis can often take weeks to months (78–80) and biofilms can generally form in a matter of hours to days, mechanisms must be present to prevent microbial growth on externally laid eggs.

Many female cephalopods maintain a symbiotic bacterial community within their reproductive system in an organ known as the accessory nidamental gland [ANG, Fig. 1A-B, (65)]. The ANG community is composed predominantly of *Alphaproteobacteria*, *Gammaproteobacteria*, and *Verrucomicrobia*, depending on the cephalopod species (68, 69, 71–73, 239). These bacteria are added to the egg jelly coat (JC, Fig. 1C) prior to clutch deposition on the substrate, after which the bacterial community is typically stable through embryogenesis (72). As the eggs do not receive parental protection, they face potential threats from fouling microorganisms and/or predation. Many hypotheses have been proposed for the functional role of the ANG bacteria, including assisting in sexual maturation of the host (75), contributing to formation of the egg capsule (65, 240), or providing protection for the developing embryos

against pathogenic or fouling organisms (70). However, while the antibacterial activity of a few specific ANG/JC bacterial isolates and the whole ANG have been implicated (85, 241), the function of this association remains uncharacterized.

Demonstrating function in host-microbe associations can prove difficult, necessitating tractable experimental models and diverse approaches (242). The Hawaiian bobtail squid, *Euprymna scolopes,* offers just such an experimental model for studying symbiotic interactions. *E. scolopes* is well-known for its association with the light organ symbiont *Vibrio fischeri*, and more recently for its ANG consortium (72, 87). The hosts are easily maintained in the laboratory with minimal effect on the composition of the ANG bacterial community (69, 72), which is dominated by *Alphaproteobacteria* from the *Rhodobacteraceae*, mostly *Leisingera* spp., as well as *Verrucomicrobia* (69, 72, 239). The ease of maintaining this squid in the laboratory and the stability of its ANG community in captivity and throughout embryogenesis (72) makes this system ideal for investigating the function of ANG bacteria in cephalopods.

Here we present evidence for the role of the ANG bacterial consortium in cephalopod reproduction, integrating knowledge of community composition, functional ecology, chemistry, and natural product discovery to begin to understand the role of bacterial members within the symbiosis. Using antibiotic treatment and egg manipulation, we demonstrate that ANG bacteria from *E. scolopes* provide resistance to fouling fungi. In addition, the majority of bacterial strains isolated from the ANG or eggs, as well as extracts from those strains, inhibited the fungus *Fusarium keratoplasticum* and/or the yeast *Candida albicans*. Microbial interaction-driven molecular network analyses revealed potential compounds that may be involved in preventing fungal overgrowth, which we used to identify several bioactive secondary metabolites.

Experimental Procedures

Antibiotic Clutch Experiments

Adult female *Euprymna scolopes* were collected from a protected sandflat in Maunalua Bay, Oahu, HI (21°26'3.36" N, 157°47'20.78" W) and were shipped to the University of Connecticut to be maintained in aquaria. Egg clutches laid in captivity by these squid were split into similarly sized groups and placed in aerated filter-sterilized artificial seawater (FSASW). One group was treated with an antibiotic cocktail (25 µg/mL each of penicillin G, kanamycin, spectinomycin, streptomycin, and gentamicin, n=5) and monitored over a four-week period with daily water changes with fresh antibiotics added. A second group from the same clutch was untreated and maintained under the same conditions without the addition of the antibiotic cocktail (n=5). Clutches were maintained in a water bath at 25-27 °C for the duration of experiments. The viability of each group was determined once hatching was complete by dividing the number of unhatched eggs by the number of total eggs (number of unhatched eggs added to the number of juveniles hatched). The health of a subset of hatched juveniles was checked by inoculating the water with the light organ symbiont, *Vibrio fischeri*, and monitoring colonization using previously described methods (224). Percent hatch was compared between untreated and antibiotic-treated clutches using a paired two-tailed t-test.

The effect of antibiotic cocktail on embryo health was assessed by placing eggs under conditions that would not promote the growth of fungi. Clutches (n=3 for each group) were placed in a laminar flow hood with 0.22 µm FSASW and covered for the duration of the experiment. All equipment was UV sterilized for a minimum of one hour prior to the addition of eggs. Clutches were aerated with pumps placed inside the laminar flow hood to prevent airborne contaminants from entering the water. FSASW was changed and fresh antibiotic cocktail was

added every 2-3 days. Associated untreated controls from the same original clutches received no antibiotic cocktail. In a separate experiment, groups of clutches with and without antibiotic treatment were maintained under lower temperatures (15-20 °C) to prevent fungal growth (n=4). Neither growth nor germination of *F. keratoplasticum* FSSC-2g were observed under these conditions. The viability of each clutch was determined as described above.

To test antibiotic treatment under more natural conditions, clutches deposited by a wild-caught female bobtail squid were maintained in a tank with running unfiltered Hawaiian seawater and sand at Kewalo Marine Laboratory, Oahu, HI. These clutches (n=3) were divided into two groups; treatment with 25 µg/mL of antibiotic cocktail and an untreated control. Clutches were maintained for one week in an aerated beaker with daily water changes and fresh antibiotics in the treatment group. After one week, both groups were transferred back to a tank with running seawater for the remainder of embryogenesis. The untreated group had 94% viability, while the antibiotic-treated section had 10% clutch viability and developed heavy fouling.

In order to test the effects of treatment using a different antibiotic, clutches were either treated with chloramphenicol (20-25 µg/mL), left untreated, or treated with ethanol or water only (solvent controls for chloramphenicol). Laminar flow hood experiments were repeated as described above. Because chloramphenicol was determined to have a negative impact on hatch rates (Figure S2), all subsequent experiments were conducted with the antibiotic cocktail.

Fungal Isolation and Characterization. *Fusarium keratoplasticum* FSSC-2g (CT 12-1807) was originally isolated from antibiotic treatment experiments as described above. *F. keratoplasticum* FSSC-2i (HI 66667) was isolated from experiment conducted in natural Hawaiian seawater as

described above. The outer capsule of an egg that did not hatch and a decapsulated egg containing a viable embryo from the antibiotic-treated section were both plated onto seawater tryptone [SWT, (91)] to sample fungi.

Fungal cultures were further isolated on inhibitory mold agar with 0.05 g/L gentamicin. Initial identification of fungal isolates occurred by analysis of morphologic characteristics on carnation leaf agar and DNA sequence analysis of the ITS-1 region of rRNA. PCR was performed using two general fungal ITS-1 primers (ITS-5, 5'-GGAAGTAAAAGTCGTAACAAGG-3'; ITS-2, 5'-GGAAGTAAAAGTCGTAACAAGG-3', 41) and GoTaq polymerase (Promega, Madison, WI) with the following profile conditions: initial denaturation for 3 min at 95 °C; 30 cycles of denaturation for 30 sec at 95 °C, annealing for 30 sec at 62 °C, and elongation for 1 min at 72 °C. PCR products were sequenced in both directions and identified as *F. keratoplasticum* via databases BLASTn analysis at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) and RDP (http://rdp.cme.msu.edu/seqmatch).

Portions of three loci (*TEF1*, *RPB2*, ITS + LSU rDNA) were used to identify three *Fusarium* strains used in the present study as *F. keratoplasticum* following a published protocol (243). Arabic numerals and lowercase roman letters, respectively, identify species and unique 3-locus haplotypes (Fig. S3). Maximum parsimony (MP) and maximum likelihood (ML) phylogenetic analyses were conducted, respectively, with PAUP* 4.0b10 (244) and GARLI 2.01 (245). MP analyses were conducted using the heuristic search option, the tree bisection–reconnection (TBR) branch-swapping algorithm with MULPARS on and 1000 random sequence addition replicates. The CIPRES Science Gateway TeraGrid (https://www.phylo.org/) (246) was used to conduct the ML analysis employing the GTR + I + Γ model of molecular evolution. Clade support was assessed by 1000 MP and ML bootstrap pseudoreplicates of the data.

Egg Component Experiments. A method developed for harvesting conidia of *Fusarium oxysporum* (247) was modified for use with *F. keratoplasticum*. *Fusarium keratoplasticum* FSSC-2g was grown on a rotary shaker for 3-5 days in SWT supplemented with 50 μg/ml chloramphenicol at 30 °C. The culture was then strained through sterile gauze and centrifuged for 10 min at 3,200 x g. Conidia were rinsed twice with sterile water and re-pelleted after each wash followed by resuspension in one-fifth of the starting volume of sterile water. Conidia were quantified using a hemocytometer and correlated with OD_{530} to determine a conversion factor of $3x10^6$ conidia/mL/OD_{530}. Refrigeration for up to one week was found to have no effect on conidial viability.

E. scolopes eggs from one clutch were dissected using sterile forceps into three groups: intact eggs, eggs lacking outer capsule, and eggs lacking outer capsule and jelly coat (n=4 separate trials; 8-10 eggs/group). Initial dissections were completed at day 5 of embryogenesis, as earlier dissection prevented embryo development. Each group was challenged with *F. keratoplasticum* FSSC-2g, 10^4 conidia/mL over the course of embryogenesis, which is approximately an order of magnitude higher than eggs would be likely to encounter naturally (238). Eggs were maintained in FSASW with aeration, and water changes with fresh conidia were conducted every 2-3 days. For two of the four trials, two eggs from each group were sampled and monitored every 3-4 days for the presence of fungal hyphae.

Transmission Electron Microscopy. Fungal fouled eggs that had been treated with antibiotics were prepared for TEM by first decapsulating the eggs and then fixing and embedding using established protocols [n=3, (9,10)]. Control eggs were not exposed to antibiotics (n=3). Briefly,

eggs were fixed (2.5% gluteraldehyde/2% paraformaldehyde solution) and stained with a solution of 1% osmium – 0.8% potassium ferricyanide. Eggs were embedded in Spurr's epoxy resin and sectioned on a Leica UCT Ultramicrotome (Leica Microsystems, Buffalo Grove, IL) into 90 nm ultrathin sections. Samples were imaged on an FEI Technai Biotwin transmission electron microscope (FEI, Hillsboro, OR).

Antibiotic-treatment Effects on Bacterial Communities. Antibiotic-treated clutches were maintained in a laminar flow hood to prevent infection as described above. FSASW was changed every two-three days and fresh antibiotics were added. Five eggs from the same clutch (n=5 clutches) were sampled on days 0, 10, and 15 of embryogenesis. JCs were isolated by first surface sterilizing in ethanol, followed by homogenization in filter-sterilized squid Ringer's solution [FSSR, (69)], and then plated on SWT to quantify culturable JC bacterial abundance. Differences in abundance between antibiotic-treated JCs and untreated JCs were analyzed via paired two-tailed t-tests for each time point (Fig. 1L).

The fungal biomasses that formed on clutches treated with antibiotics were homogenized in FSSR (n=12) and bacterial DNA from these samples was extracted using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. DNA concentration was determined using the Qubit® dsDNA High Sensitivity assay (ThermoFisher Scientific Inc., Waltham, MA) and averaged 1.13 ± 0.79 ng/µl. The V4 region of the 16S rRNA gene was sequenced on an Illumina MiSeq (Illumina, San Diego, CA, USA) following established protocols (72)(248)(249). Sequence data were analyzed using QIIME (250) and NMDS plots of Bray Curtis beta-diversity analyses were created in R using the VEGAN package (251) as previously described (72). Sequence data were compared to ANG/JC community data

previously published under the project ID ENA PRJEB14655, accession numbers ERS1498392 to ERS1498398, ERS1496666 to ERS1496676, and ERS1496678 to ERS1496722 (72). Bacterial sequences from the fungal biomasses were deposited in the European Nucleotide Archive (ENA) under the project ID PRJEB23346.

Well Diffusion Assay. Bacterial isolates from the ANG and JC were obtained from previous studies (69)(103)(85) and this study (Table S1) by homogenization in FSSR and plating serial dilutions on SWT and R2A media. Isolates from this study were identified to the genus level by Sanger sequencing of the 16S rRNA gene using universal primers [27F and 1492R, (119)], and BLASTn search of the 16S rRNA NCBI database.

ANG/JC isolates were assayed for their ability to inhibit *F. keratoplasticum* FSSC-2g on SWT or SWT without glycerol (SWT_{ng}) agar plates using an adaptation of the plate diffusion assay described by Fraune et al. (235). SWT_{ng} media was tested since glycerol can inhibit secondary metabolite production in some bacteria (252). Isolates were grown in 3.0 mL SWT or SWT_{ng} broth at 30 °C, shaking at 120 rpm, until cultures reached stationary phase as measured by OD_{600}. A sterile 6.0 mm diameter borer was used to make wells in the agar and 60 µl of each isolate culture was added to wells on separate plates. These plates were incubated for 48 hours at 26 °C to allow the isolate to form a biofilm around the walls of the well. *F. keratoplasticum* was then added to the wells at 10^4 conidia/mL and plates were further incubated under the same conditions for an additional four days. After incubation, plates were photographed and the area of visible hyphal growth was measured using the image analysis software program FIJI (120). Cyclohexamide (1000 µg/mL) was used as a positive control and SWT or SWT_{ng} broth (60 µl) was used as a negative control. Sterile water, which was used to dilute the *F. keratoplasticum*

conidial stock in this assay, was also used as an additional negative control. For each bacterial strain, assays were performed using three separate experimental replicates with three technical replicates per experiment. Inhibition was reported as percent of hyphal growth compared to hyphal growth in the broth negative control. The total number of isolates which fell into each inhibition category (Fig. S5b) was determined from both SWT and SWT_{ng} assay results. If degree of inhibition varied between media type, each isolate was only counted once and included in the stronger inhibition category.

Culture and Extraction of Microbial Isolates. ANG and JC bacterial cultures were prepared for extraction using a modification of the previously described three step culture protocol (85). All bacteria were cultured in SWT media. Small-scale cultures were prepared by inoculating 2-3 bacterial colonies into 5 mL of media, medium scale cultures utilized 1.5 mL of each small-scale culture into 50 mL of media, and large-scale cultures involved transfer of 15 mL of medium scale cultures into 500 mL of media. All cultures were incubated for 3 days at 28 °C while shaking at 200 rpm.

Large-scale bacterial cultures were extracted as previously described (85). Briefly, pre-washed Diaion HP20 resin (50 g, 10 % w/v; Supelco, Bellefonte, PA, USA) was added to sonicated cultures and allowed to incubate for 24 h at room temperature while shaking (125 rpm). The resin and bacterial culture were then filtered, washed with water (discarded), and extracted with methanol (MeOH), dichloromethane (DCM), and acetone (2 x 150 mL; ACS grade, Sigma Aldrich, St. Louis, MO, USA). Combined organic extracts were concentrated *in vacuo* and partitioned with ethyl acetate and water to remove residual aqueous material.

Fusarium keratoplasticum FSSC-2g was also cultured and extracted for use in metabolomics experiments. *F. keratoplasticum* FSSC-2g was grown on SWT agar as described below and transferred into 50 mL of fresh SWT media. Pre-washed Diaion HP20 resin (5 g, 10 % w/v) was added and incubated and extracted as described above, using 2 x 30 mL of each solvent.

96-well Liquid Antifungal Assays. Bacterial extracts were tested for antifungal activity against *F. keratoplasticum* FSSC-2i, FSSC-2g, and FSSC-2d. Preparation of fungal inoculum followed previously described methodology (253) with each fungal isolate grown on SWT agar supplemented with 50 μg/mL of chloramphenicol at 28 °C. Each culture was covered with 3 mL of media, probed with a loop, and transferred into 10 mL of fresh media. The fungal mixture was vortexed, allowed to settle, and the cell suspension below hyphae was transferred to a new Falcon tube. The cell suspension was adjusted to OD_{600} 0.15-0.17 and diluted 1:50 with media to obtain a working solution.

F. keratoplasticum assays were performed according to the Clinical and Laboratory Standards Institute (CLSI) M38-A2 (254) in 96-well flat bottom plates (Corning Costar™, Kennebunk, ME, USA) with the following modifications. Each well consisted of 100 μL of diluted fungal inoculum, 98 μL of media, and 2 μL of either positive control (amphotericin B, final testing concentration 4 μg/mL), dimethyl sulfoxide (DMSO) negative control, or extract prepared in DMSO (final testing concentration 500 μg/mL). Extracts and controls were tested in two separate experimental replicates with three technical replicates per experiment. Plates were read on a Synergy H1 Hybrid Reader (Biotek, Winooski, VT, USA) at 600 nm at 0 h, incubated

at 28 °C, and read again at 14 h. Percent growth was calculated in comparison with DMSO, the negative control.

Bacterial extracts were also tested against the human fungal pathogen *Candida albicans* ATCC 18804 following CLSI M27-A2 (254) in 96-well flat bottom plates with the following modifications. Yeast colonies were selected from an agar plate and inoculated into 5 mL of RPMI media (Gibco®) supplemented with 2% glucose, and then adjusted to OD_{600} 0.08-0.1 and diluted 1:10 with fresh media to obtain a working solution. Each well consisted of 100 μL of diluted fungal inoculum, 98 μL of media, and 2 μL of either positive control (amphotericin B, final testing concentration 2.5 μg/mL), DMSO negative control, or extract prepared in DMSO (final testing concentration 500 μg/mL). Extracts and controls were tested in two separate experimental replicates with three technical replicates per experiment. Plates were read on a Synergy H1 Hybrid Reader at 600 nm at 0 h, incubated at 35 °C, and read again at 24 h.

Fungal Challenge of Intact Egg Clutches. Egg clutches were divided into two portions, one of which served as an untreated control and the other challenged by 10^4 conidia/mL *F. keratoplasticum* FSSC-2g for 11 days in FSASW at room temperature with water changes and addition of fresh conidia every two-three days (n=4, of which one set of clutches was used for networking in Fig. 6). Experiments did not involve antibacterial treatment, leaving any JC-associated bacteria intact and potentially capable of producing defensive metabolites. Challenged and control eggs (5 mL volume each) were separately extracted with 2:1 DCM:MeOH (7 x 50 mL). Samples were sonicated, manually macerated with a metal spatula, and incubated for

10 min at room temperature, after which solvent was filtered, combined, and concentrated *in vacuo*.

LC-MS/MS Analyses and Molecular Networking. Extracts from all clutches were analyzed via LC-MS/MS using a Waters Synapt G2-Si HDMS mass spectrometer coupled to a Waters Acquity ultra-high performance (UPLC) liquid chromatography system. Separations utilized a Waters Acquity UPLC HSS T3 column (2.1 x 150 mm, 1.8 μM, 100Å) connected to an Acquity UPLC HSS T3 VanGuard pre-column (2.1 x 5 mm, 1.8 μM, 100Å). Samples were prepared at 1 mg/mL in a 1:1 mixture of methanol/water and analytes were chromatographically separated using a linear 0.45 mL/min gradient with the following conditions: 0.5 min hold at 95% A (0.1% formic acid in water) and 5% B (0.1% formic acid in acetonitrile), 3.5 min ramp to 60% B, 4 min ramp to 98% B, 1 min hold at 98% B, 0.2 min ramp back to 5% B and a re-equilibration hold at 5% B for 1.8 min.

Data were obtained using an 11 min Fast data dependent acquisition (Fast DDA) method with a 0.1 sec MS survey scan ranging from *m/z* 50-2000, followed by five 0.1 sec data-dependent MS/MS scans of ions with intensities greater than 1000 as determined during the preceding scan event. Survey scans were acquired without collision energy (CE) and MS/MS scans were acquired using a linear two-stage low mass (0-20V) and high mass (10-40V) CE linear ramp. The instrument was operated in positive, resolution mode using the following parameters: 2 kV capillary voltage, 100 °C source temperature, 20V sampling cone, 800 L/hr desolvation gas flow, and 80V source offset. Lockspray real-time mass correction was turned on to achieve optimal mass accuracy by monitoring 556.2771 m/z and 120.0813 m/z ions from

simultaneous infusion of a 400 pg/μL leucine enkephalin solution (0.1% formic acid in a 50% water, 50% methanol solution) with 0.1 sec scans, a 10 sec scan interval and 3 average scans per correction.

Data were converted to mzML format using ProteoWizard package (255) and processed in GNPS: Global Natural Products Social Molecular Networking [http://gnps.ucsd.edu, (256)]. For the challenged clutch experiment, GNPS parameters were set to: minimum cosine similarity score of 0.65 with at least four matching peaks; parent mass tolerance of 1.0 Da; fragment ion tolerance 0.02 Da; and filter peaks in a 50 Da window was turned off. The first network with ANG and JC bacterial extracts utilized the same parameters with the exception of having a fragment ion tolerance of 0.3 Da. Active extracts (0-25% growth of at least one *Fusarium* spp.) were grouped for these analyses and included *Labrenzia* sp. ANG18, *Leisingera* sp. ANG-S2, *Leisingera* sp. ANG15, *Ruegeria* sp. ANG-R, *Ruegeria* sp. ANG10, *Alteromonas* sp. JC21, *Pseudoalteromonas* sp. JC22, and *Vibrio* sp. JC34. Inactive extracts (≥ 76% growth of all three *Fusarium* spp.) were grouped separately from active extracts and included *Ruegeria* sp. ANG17, *Leisingera* sp. ANG-M1, *Leisingera* sp. ANG-S, *Leisingera* sp. ANG-VP, *Nautella* sp. ANG-M5, *Leisingera* sp. ANG-M7, *Leisingera* sp. ANG1, *Ruegeria* sp. ANG6, *Leisingera* sp. ANG-DT, *Leisingera* sp. ANG-S3, *Tateyamaria* sp. ANG-S1, and *Tenacibaculum* sp. JC62. Files were imported into either Cytoscape 2.8.3 (nodes arranged with FM3 layout) or Cytoscape 3.5.1 (nodes arranged using default layout) (257).

Standards of the mycinamicins (generously provided by David Sherman, University of Michigan), lincomycins (Santa Cruz Biotechnology, Inc., Dallas, TX), and lyso-PAFs (Cayman Chemical, Ann Arbor, MI) were used to determine presence in clutch extracts using a Waters Xevo G2-XS QToF mass spectrometer coupled to a Waters Acquity UPLC. Separations utilized

a Waters Acquity UPLC HSS T3 column (2.1 x 50 mm, 1.8 μM, 100Å) connected to an Acquity UPLC HSS T3 VanGuard pre-column (2.1 x 5 mm, 1.8 μM, 100Å). Conditions were identical to those described above with the exception that MS/MS scans were acquired using a linear two-stage low mass (10-40V) and high mass (40-80V) CE linear ramp. The instrument was operated in positive, sensitivity mode using the following parameters: 3 kV capillary voltage, 100 °C source temperature, and 30V sampling cone. Lockspray real-time mass correction used a simultaneous infusion of a 200 pg/μL leucine enkephalin solution. High resolution MS (HRMS) data for features of interest were used to search the METLIN database (258) for matching metabolites with a mass defect within 20 ppm.

Results and Discussion

Antibiotic treatment of eggs leads to fungal fouling.

To gain insights into the function of the JC bacteria, eggs were treated with an antibiotic cocktail (penicillin G, kanamycin, spectinomycin, streptomycin, and gentamicin, see *SI Appendix,* Experimental Procedures). The bacterial load in untreated eggs was similar to what had previously been described (72), while treated clutches experienced reduction in JC bacterial loads by 98 % throughout embryogenesis (Fig. 1L) and developed heavy biofouling (Fig. 1G-H). Untreated clutches kept under the same conditions did not develop any biofouling (Fig. 1D-E). Egg biofouling resulted in very low hatch rates, averaging 9 %, compared to 58 % for untreated clutches (t_4=3.572, p=0.023, Fig. 2). The biomass was dominated by fungal hyphae interspersed with bacterial cells (Fig. 1J-K). Fungal fouling visible by eye typically appeared on antibiotic-treated clutches between days 8-14 of embryogenesis (n=17 clutches, Fig. 1G-H). However,

fungal hyphae appeared on the antibiotic-treated egg surface as early as day 3 of embryogenesis (Fig. S1B) and developed into heavy biofouling by day 10 (Fig. S1H). The expanding fouling completely enveloped the clutch by day 15 (Fig. S1I-J). In contrast, no fungal hyphae appeared on untreated eggs from the same clutch at any point during development (Fig. S1E-G, K-M). Microconidia and hyphae were found dispersed throughout the JC layers but did not penetrate the chorion of antibiotic-treated eggs (Fig. 1I). The antibiotic-treated JCs also contained fewer bacterial cells compared to the JCs of untreated eggs (Fig. 1F, 1I, 1L). When eggs were treated with the antibiotic cocktail at a lower temperature of 15-20 °C (to inhibit fungal growth), or under sterile conditions, hatching success was similar to untreated eggs (Fig. 2). The resulting juveniles were also successfully colonized with the light organ symbiont, *Vibrio fischeri*. Thus, treatment with the antibiotic cocktail did not affect embryonic development or post-hatching bacterial light organ colonization. Treatment with chloramphenicol alone (not present in the cocktail) led to the formation of a similar fungal biomass as observed with treatment with the antibiotic cocktail, but negatively influenced hatching success (Fig. S2).

Fungal isolates from the fouled clutches were analyzed via multilocus sequence typing (MLST) and determined to be haplotype 2g of *Fusarium keratoplasticum*, a recently described species nested in the *Fusarium solani* species complex [(FSSC), Fig. S3, (259)]. *Fusarium keratoplasticum,* together with at least 20 species in the FSSC, has been implicated in opportunistic infections of humans and other animals (243, 260–262). Many of these persistent infections are difficult to treat because members of the FSSC and other fusaria are broadly resistant to currently available antifungal drugs (263).

Eggs treated with a pulse of antibiotic cocktail and maintained in flowing Hawaiian seawater also developed fungal fouling dominated by *F. keratoplasticum*, demonstrating that this

fungus is present and capable of fouling clutches in conditions similar to the host's natural habitat. Four fungal isolates from experiments conducted using Hawaiian seawater were identified as *F. keratoplasticum* haplotype FSSC-2i (Fig. S3).

Profiling of the bacterial community of the fouling biomass that formed on the eggs during antibiotic treatment revealed that it was dominated by *Gammaproteobacteria* and *Alphaproteobacteria*, along with nine other bacterial classes (Fig. S4A). These fouling-associated bacteria were distinct from a previous analysis of the ANG/JC communities [Fig. S4B, (72)], and the bacterial community composition also varied widely between samples (Fig. S4C). Therefore, these fouling-associated bacteria likely consist of environmental bacteria for which the fungal biomass provided a suitable substrate, but some of these may also be opportunistic pathogens of the eggs.

Symbiont-containing egg JC is protective against fungal fouling.

To discern the functions of the bacterial community from potential embryonic factors that may have a protective effect, eggs were dissected into separate components and challenged with a conidial suspension of *F. keratoplasticum* FSSC-2g at 10^4 conidia/mL [approximately an order of magnitude higher concentration than probable environmental levels, (238)] (Fig. 3B,D,F). These eggs were not treated with antibiotics, leaving the egg bacterial community intact. By day 18 of embryogenesis, when the embryos were mature and near hatching, challenged whole eggs showed negligible signs of fungal infection (Fig. 3B), with only a few hyphae noted on the surface of these eggs and no apparent effect on the development of the embryos. Similar results were observed when the outer capsule was removed, leaving the JC and its bacterial community intact (Fig. 3D). However, when both the outer capsule and the JC were removed, leaving only

the developing embryo within its yolk sac, eggs were heavily fouled by a fungal biomass and the embryos were not viable (Fig. 3F). Control eggs developed normally and were not fouled (Fig. 3A, C, E). These experiments, along with the antibiotic treatments discussed above, suggest that the JC and its bacterial community provide protection against *F. keratoplasticum*.

Symbiotic bacteria and their metabolites inhibit fungi.

To better understand the potential antifungal roles of specific bacterial strains that comprise the ANG and JC bacterial communities, 32 of 33 bacterial isolates were tested using a well diffusion inhibition assay against *F. keratoplasticum* FSSC-2g [one strain, *Ruegeria* ANG-S4, grew very slowly in culture and so was excluded from this assay but was included for extract assays (see below)]. Three *Alphaproteobacteria* strains (*Leisingera* spp. ANG-S and ANG15, and *Labrenzia* sp. ANG18) and four *Gammaproteobacteria* strains (*Alteromonas* sp. JC21, *Vibrio* sp. JC34, and *Pseudoalteromonas* spp. JC22 and JC28) were strong inhibitors of *F. keratoplasticum* FSSC-2g, reducing the area of hyphal growth to 0-25 % of the negative control (Fig. S5, Table S2). Ten strains showed moderate inhibition (26-50 % growth), eleven strains showed weak inhibition (51-75 % growth) and four strains exhibited minimal or no inhibition (≥76 % growth, Fig. S5, Table S2). Together, our results show that 28 out of 32 (87.5 %) ANG/JC strains have the physiological potential to inhibit *F. keratoplasticum*.

To investigate the role of bacterial secondary metabolites in protecting the eggs from fungal fouling, 33 bacterial isolates from the ANG and JC were cultured, extracted, and tested in a 96-well liquid assay against three *F. keratoplasticum* strains: FSSC-2i (Hawaiian isolate), FSSC-2g (Connecticut lab isolate), and FSSC-2d (human clinical strain). In addition, extracts were tested against the yeast, *Candida albicans* ATCC 18804, a common member of the human

microbiota that can be a significant opportunistic pathogen. Overall, eight of the 33 extracts exhibited strong inhibition (0-25 % growth) against at least one *F. keratoplasticum* strain, an additional nine extracts exhibited moderate inhibition (26-50 % growth), and five extracts showed weak inhibition (51-75 % growth, Fig. 4). In total, 72.7 % of the ANG and JC bacterial extracts showed antifungal activity, suggesting a strong functional role of these symbionts in protecting embryos against fungal infection.

While six bacterial extracts (four from *Alphaproteobacteria* and two from *Gammaproteobacteria)* exhibited moderate to strong inhibition against all three strains of *F. keratoplasticum*, several extracts inhibited only one or two of the three strains while not inhibiting the other(s) (Fig. 4C). A greater number of bacterial extracts were classified as strongly or moderately active against FSSC-2i/2g in comparison to FSSC-2d (Fig. 4B). Given that FSSC-2d is a clinical strain and not isolated from squid eggs, this differential activity may be due to adaptation of active metabolites to different fungal strains. Ten ANG and JC bacterial extracts (30.3 %) also exhibited strong or moderate inhibition of *C. albicans* (Fig. 4A). The ANG and JC extracts often showed differential activity with some only active against *F. keratoplasticum* (e.g., *Ruegeria* spp. ANG10 and ANG-R) while others were only active against *C. albicans* (e.g., *Leisingera* spp. ANG1 and ANG-M7, Fig. 4A). Future research will focus on expanding these assays to testing other fungi to determine whether JC symbionts differentially inhibit a variety of fungi.

Overall, seven of the 32 isolates (21.9 %) and eight of the 33 extracts (24.2 %) exhibited strong inhibition against *F. keratoplasticum*. In other systems, such as the cnidarian hydra, interactions between multiple members of the epithelial bacterial community are necessary to protect the host from fusarial infections (235). Synergistic antifungal activity has not yet been

determined in the bobtail squid system but similar effects may be present given the complexity of the bacterial community found in both the ANG and JC.

Our experimental data demonstrate that ANG and JC bacteria, and associated organic extracts from these strains, inhibit fungi. Bacterial secondary metabolite protection from fungi has been documented in several invertebrate systems. In the Oriental shrimp, *Palaemon macrodactylus*, and in the American lobster, *Homarus americanus*, eggs are associated with protective bacteria that produce the antifungal compounds 2,3-indolinedione (isatin) and 4-hydroxyphenethyl alcohol (tyrosol) respectively (54, 234). When *Hydra vulgaris* was treated with antibiotics to remove the resident epithelial microbiota, the animals developed a detrimental infection of unidentified *Fusarium*, which was rendered ineffective by then complementing the treated hosts with members of their original bacterial community (235). Female beewolves apply *Streptomyces philanthi* symbionts from antennal reservoirs to the brood chambers of their larvae, which are then incorporated into the larval cocoon to protect against environmental fungi. This protection is largely accomplished by the bacterial symbionts' production of piericidin, streptochlorin, and nigericin derivatives (6, 46, 48). These examples demonstrate that diverse aquatic and terrestrial invertebrates utilize symbiotic bacteria to ward off fungal infections, and our work provides additional evidence to support a similar functional role of ANG and JC bacteria in *E. scolopes*.

Chemical networking analyses reveal potential antimicrobial metabolites.

To further understand the function of ANG- and JC-associated bacteria, we conducted experiments to identify antimicrobial secondary metabolites from ANG and JC bacterial isolates.

Utilizing a unique combination of LC-MS/MS analyses with biological activity of ANG and JC bacterial extracts (Fig. 4), we created a microbial interaction-driven molecular network [Fig. 5A, (264)] to compare compounds from bacterial extracts found to strongly inhibit at least one *F. keratoplasticum* strain (grouped as active) or ones that failed to inhibit any of the *F. keratoplasticum* strains (grouped as inactive) with compounds also found in egg clutches, including both untreated clutches as well as clutches challenged with *F. keratoplasticum* (experimental design shown in Fig. 6A and discussed below). This network analysis allowed for the identification of metabolites found in egg clutches that were also present in antifungal bacterial extracts (Fig. 5). As metabolites found in egg clutches are hypothesized to be derived from associated bacteria, mass spectral features were prioritized when co-occurring in antifungal bacterial extracts and either the challenged or control clutches (Fig. 5B), resulting in identification of 10 features of interest. One feature, m/z 393.1480, which was present in the active extract of *Labrenzia* sp. ANG18 and the control clutches, was determined to be lincomycin B (Fig. 5C), a structural analogue of the FDA-approved antimicrobial drug, lincomycin A. Upon subsequent examination of the raw LC-MS/MS data, lincomycin A was also present in the control clutches but not detected in the *Labrenzia* sp. ANG18 isolate. These results were confirmed via comparison with commercial standards (Fig. S6).

The lincomycins are the founding members of the lincosamide class of bacterial natural products first isolated in 1962 from the soil-derived actinomycete *Streptomyces lincolnensis* var. *lincolnensis* (265). Lincomycins have potent activity against Gram-positive bacteria (266), with lincomycin A (marketed as Lincocin®) receiving FDA approval in 1967. Biosynthesis of the lincomycins involves unusual combinations of nonribosomal peptide synthetase (NRPS) and amino sugar enzymes to attach an *N*-methylated 4-propyl-L-proline to a methylmercapto eight-

carbon sugar (267, 268). Commercially available lincomycins A and B were tested herein for antifungal activity against *F. keratoplasticum* FSSC-2i and -2g and found to be inactive, however several additional lincomycin derivatives were preliminarily identified from this *Labrenzia* strain and future experiments will focus on isolating these derivatives and testing them for antifusarial activity. In addition, although these metabolites were inactive against *F. keratoplasticum*, community sequencing of the fouling biomass also detected the presence of bacteria (Fig. S4), and thus the presence of lincomycins A and B in egg clutches may indicate a general antimicrobial response from the JC community. Regardless, the presence of this class of molecules in the egg clutches and in strain *Labrenzia* sp. ANG18 provides evidence for the functional potential of the ANG and associated bacteria in protecting squid eggs from infection.

In continued exploration of the metabolites involved in the antimicrobial protective effects of ANG- and JC-associated bacteria, we conducted an experiment to identify compounds induced in the presence of *F. keratoplasticum* (Fig. 6). Egg clutches were divided into two portions, with one part utilized as an untreated control while the other was challenged with *F. keratoplasticum* FSSC-2g (Fig. 6A). These experiments did not involve antibiotic treatment, leaving JC-associated bacteria intact and potentially capable of producing defensive metabolites. Extracts from these two clutch portions as well as from a culture of *F. keratoplasticum* FSSC-2g (used to control for fungal-derived metabolites in the network) were all networked (Fig. 6B), with several mass spectral features found only in *F. keratoplasticum*-challenged clutches, including two highlighted clusters of metabolites (Fig. 6C). Further analyses of the larger highlighted cluster revealed the presence of molecules that share several structural characteristics in common with the known antimicrobials, mycinamicins III, IV, and VI [Fig. S7; (269–271)], including high resolution MS within acceptable limits and a distinctive in-source fragment

matching the desosamine sugar moiety. Further testing of authenticated standards found that mycinamicins IV and VI exhibited weak antifungal activity against *F. keratoplasticum* FSSC-2g (73.4 % and 68.6 % growth, respectively), while mycinamicin III was inactive. Additional analyses of authenticated standards and egg samples by LC-MS/MS revealed a difference in retention times that would suggest these features were more similar to glycerophosphocholines, all known members of which had higher mass defects than the original assignments. Efforts are ongoing to unambiguously identify these features, including exploring possible effects of the gelatinous nature of the JC microenvironment on the physicochemical parameters (e.g., retention time). Regardless of their identification, production of the features associated with this cluster increased when eggs were challenged with *F. keratoplasticum,* suggesting an induced response by the JC community to fungal challenge.

Molecules from a second cluster induced in the presence of *F. keratoplasticum* (Fig. 6C), were putatively identified as glycerophosphocholines, similar to lyso-platelet activating factor (PAF) C:16 and lyso-PAF C:18. Commercial standards of these compounds were tested for activity against *F. keratoplasticum* FSSC-2i and -2g. Both compounds demonstrated antifungal activity with MIC values of 125 µg/mL. In subsequent LC-MS/MS analyses of additional clutches, we detected lyso-PAFs C:16 and C:18 in both challenged and control clutches. Further examination of the raw LC-MS/MS data confirmed that the original nodes were consistent with glycerophosphocholines, but differ slightly from lyso-PAFs C:16 and C:18, and thus may represent related molecules from this large family of lipophilic compounds. In addition, other structurally distinct glycerophosphocholine derivatives were found in the network (Fig. 6D), although this cluster contains metabolites found in the challenged clutches, the control clutches, and/or the *F. keratoplasticum* extract (Table S4). LC-MS/MS analyses of egg clutches from three

additional experimental replicates confirmed the presence of the mycinamicin-like and glycerophosphocholine features in other clutches, with overall higher mean ion intensities in challenged clutches (Table S5).

Previous reports of glycerophosphocholines, and specifically PAFs, indicate these compounds are important defensive phospholipid signaling molecules that are deacetylated by acetylhydrolases to obtain lyso-PAFs (272, 273). These interconversions have been shown to be important for non-self recognition, including antimicrobial defense in other marine host-microbe systems (274). Previous research suggests that bacteria capable of producing choline-containing molecules, such as lyso-PAFs, are usually closely associated (e.g., symbiotic) with a eukaryotic host that supplies choline for the biosynthesis of these compounds (275, 276). The JC does contain membrane-like structures that separate each layer of the egg (72), although whether these structures contain cholines is not known. While our current studies do not reveal the origin of these glycerophosphocholines, a specific suite of these molecules do appear to be induced upon fungal challenge and may play an important role in egg defense.

The detection of antimicrobial compounds in both unchallenged (lincomycins) and challenged eggs (mycinamicin-like and/or glycerophosphocholines), suggests that eggs contain both constitutive and induced metabolites, thus having a broad potential to mount challenges to fouling and pathogens. Our microbial interaction-driven network analyses revealed bioactive compounds found in both control clutches and bacterial isolates, such as lincomycins with known antibacterial activity. In addition, although identifications are still ongoing, the suite of glycerophosphocholines found only in the challenged clutches may represent a more generalized and inducible antimicrobial response of the symbiotic egg community to pathogens. Exploration of bacterial metabolites in the ANG will provide insight into whether compounds with

antimicrobial activity may be deposited into the egg from the gland itself. Despite the presence of constitutive antimicrobials in the JC, our experimental data (Fig. 2-3) suggest such compounds may not be enough to overcome fungal challenge on their own and that metabolically active bacteria are likely also required for egg defense. Finally, a number of other unidentified MS features were found in both clutches and active bacteria (Fig. 5) and these represent promising leads to discover novel bioactive compounds.

Conclusions

Our research experimentally links a defensive function to the ANG/JC symbiotic bacterial community in the protection of squid eggs in the model cephalopod *E. scolopes*. Through experimental manipulation of the egg JC and its bacteria, we demonstrate that the JC is protective against fungal fouling. We further show that multiple bacterial members of the ANG/JC are capable of inhibiting both pathogenic filamentous and yeast fungi. Furthermore, we have identified several compounds, including lincomycin, mycinamicin-like, and glycerophosphocholine derivatives, as well as yet-to-be identified compounds, that may play a role in egg defense. Although *Alphaproteobacteria* are underexplored for both natural product production and their biosynthetic gene clusters involved with these processes, microbial interaction-driven networking revealed a number of potentially exciting bioactive leads for further antimicrobial drug discovery. Because some host-microbe symbioses have evolved to select for bacterial symbionts that produce biologically active compounds (31), the ANG association may be a good source for antimicrobial drug discovery.

The ability of eggs of aquatic organisms to survive in an environment filled with potential biofoulers depends on the presence of one or more defensive mechanisms. The distribution of the

ANG symbiosis in other cephalopods suggests that a similar function occurs in other members of this group. Questions of the function of the ANG symbiosis in cephalopods have been asked since the organ was first described in *Doryteuthis pealeii* in 1909 (240, 277, 278). In this study we demonstrate that the JC bacterial community in *E. scolopes* prevents fungal fouling. Previous work has demonstrated that members of the ANG bacterial community from *E. scolopes* and the squid *Doryteuthis pealeii* can inhibit other bacteria (82, 85) and genomic analyses of ANG isolates suggests that these strains have a number of biosynthetic gene clusters that may be involved with the production of antimicrobials (103). Future work will focus on understanding whether there are conserved mechanisms of egg defense among cephalopods and whether specific pathogens are targeted. The few studies that have addressed ANG diversity suggest *Alphaproteobacteria* and *Gammaproteobacteria* are present in the ANGs of other cephalopod species (68, 71, 73, 75). In the case of *E. scolopes*, the ANG is likely colonized by bacteria from the host's environment (72). How ANG symbioses are established and maintained is another area of future research. For example, are members selected for based on their capacity to produce antimicrobial compounds and/or do other mechanisms like competitive exclusion play a role in shaping the consortium and/or egg defence? Selection of specific bacteria in binary host-microbe associations is common, but, even with complex consortia, certain bacterial taxa are often recruited [e.g., mammalian gut; (279–281)]. The mechanisms by which specific bacteria may colonize ANGs is currently unknown but a study in the market squid *D. opalescens* suggests that the morphology of the nascent ANG is poised to recruit environmental bacteria (76). In the *E. scolopes*-*V. fischeri* light organ symbiosis, partners ensure specificity through a number of mechanisms (87), and it may be that similar colonization events occur in the ANG. Our previous work has also shown that specific taxa of the consortium are segregated within tubules of the

ANG, suggesting that mechanisms are present to ensure colonization and growth of specific bacterial symbionts (69).

The phylogenetic and functional diversity of the *E. scolopes* ANG symbiotic community offers a wide range of targets for future avenues of research, including the investigation of unexplored antimicrobial activity. Future research will examine whether certain bacteria and associated extracts work synergistically against potential biofoulers and pathogens. Cultivation efforts are also underway to isolate other bacteria from the ANG/JC community, such as the *Verrucomicrobia* and *Flavobacteriia*, members of which the biosynthetic potential is unknown. In addition, a better understanding is needed of how the microenvironment of the JC (e.g., pH, jelly composition, and structural features such as the membrane-like structures; Fig. 1F) may facilitate delivery and efficacy of antimicrobial compounds. Overall, the bobtail squid-ANG association provides a tractable system to investigate functional specificity and drug discovery and to better understand the role of defensive symbioses in aquatic environments.

Acknowledgements

This research was funded by NSF IOS-1557914 to S.V.N and M.J.B. and the University of Connecticut Microbiome Research Seed Grant to M.J.B. and S.V.N. We thank the McFall-Ngai, Ruby, and Hadfield laboratories at Kewalo Marine Laboratory, University of Hawaii, for laboratory/aquarium space, Jeremy Balsbaugh and the UConn Proteomics and Metabolomic Facility for assistance in acquiring LC-MS/MS data, David Sherman and Matthew DeMars for providing mycinamicin standards, Roger Linington for access to the Natural Products Atlas, Steven Daniels and the UConn EM Facility for assistance with transmission electron

microscopy, Jessica Bertenshaw for help with the antibiotic clutch experiments, and Gail Doehring for identifying the *Fusarium* isolates using MLST data.

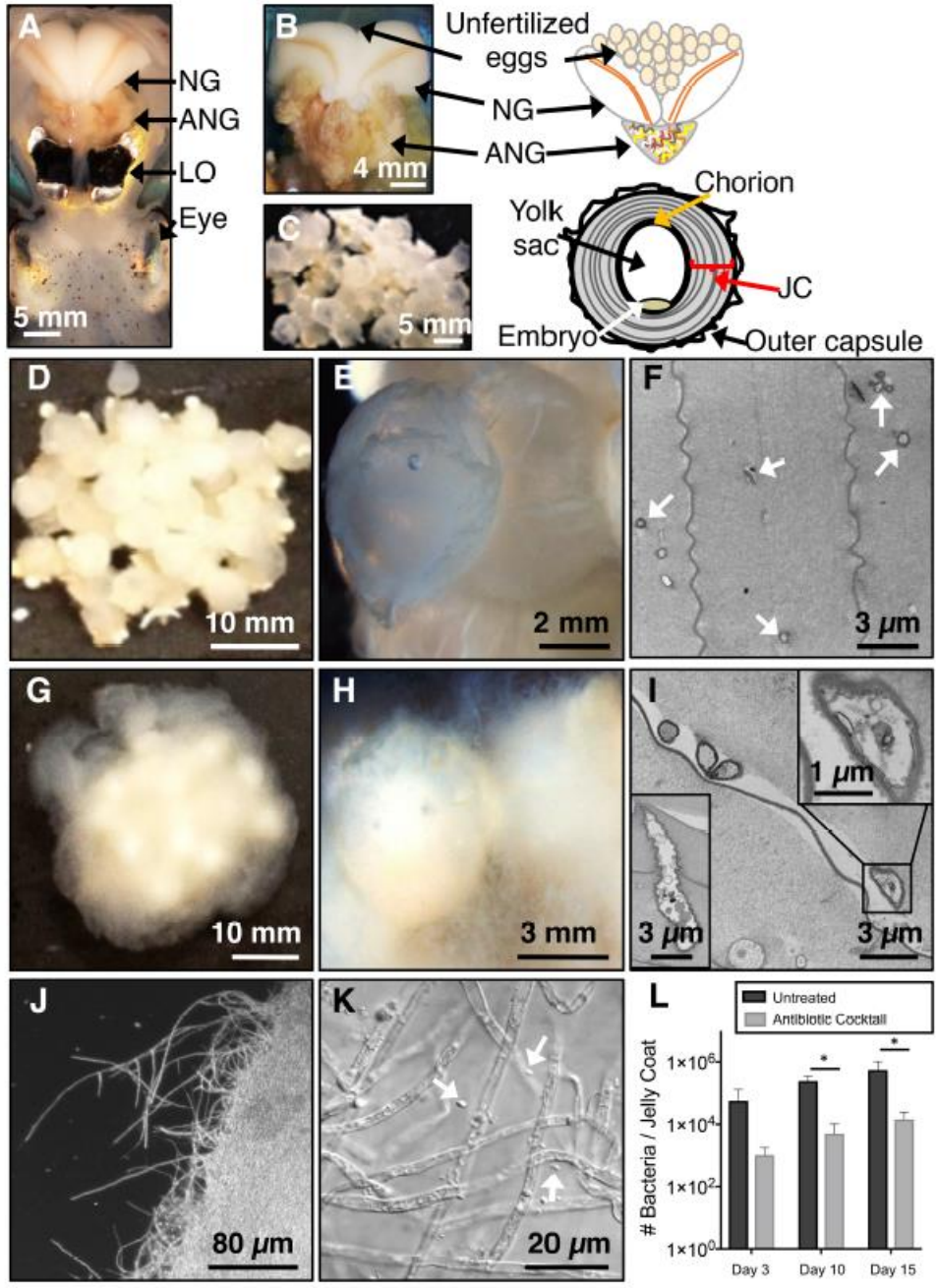
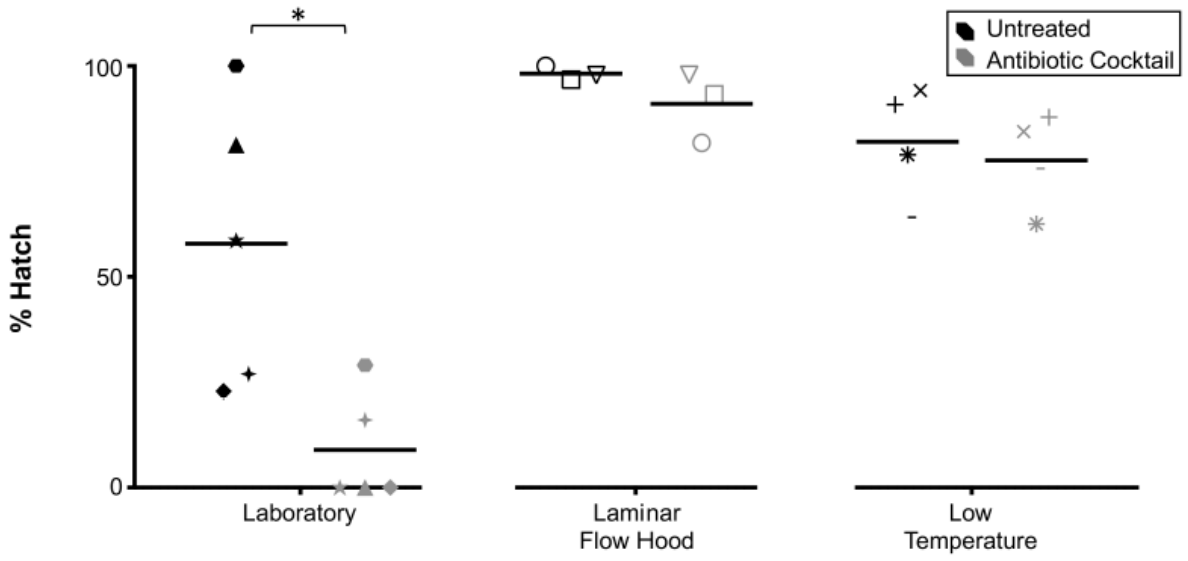


Fig. 1. Antibiotic treatment of eggs leads to fungal fouling. *(A)* Ventral dissection of a female *E. scolopes* showing location of reproductive system organs relative to the light organ and eye of the squid. Detailed image and diagram of female reproductive tract *(B)* and egg *(C)*. Treatment of clutches with an antibiotic cocktail led to the formation of fungal/bacterial fouling. Clutches

were either left untreated *(D-F)* or were treated with antibiotics (*G-I*, n=17). The fouling biomass was composed of fungal hyphae interspersed with bacterial cells (*J-K*, white arrows). Electron micrographs of JCs from eggs at Day 21 of embryogenesis show untreated egg JCs contained both single and small clusters of bacterial cells (*F*, white arrows). *(I)* The antibiotic-treated egg JCs contained numerous fungal conidia and hyphae (*I*, insets). After three days of treatment, a 98 % reduction in bacterial load was observed (*L*, t_8=1.685, p=0.131, n=5 clutches). The bacterial load of both the untreated and the antibiotic-treated clutch segments increased over embryogenesis, but a significant reduction in bacteria was found both at day 10 (t_8=5.011, p=0.001) and at day 15 (t_8=2.511, p=0.036; 98 % and 97 % respectively, graph presented as mean ± SEM). NG = nidamental gland; ANG = accessory nidamental gland; JC = jelly coat; LO = light organ.



Experiment Environment	Treatment	# of clutches	Average Eggs/Clutch	Day of Biomass Appearance	Days Hatchlings Appeared	Average % Hatch
Laboratory	Antibiotic Cocktail	5	37	8-14	17-24	9% ± 13%
	Untreated	5	27	-	17-30	58% ± 33%
Laminar flow hood	Antibiotic Cocktail	3	57	-	14-25	91% ± 8%
	Untreated	3	48	-	13-23	98% ± 2%
Low temperature (15°-20°C)	Antibiotic Cocktail	4	43	-	23-31	78% ± 11%
	Untreated	4	33	-	23-31	82% ± 14%

Fig. 2. Treatment of eggs with an antibiotic cocktail and subsequent development of fungal fouling significantly reduced hatch of juveniles. Eggs treated with an antibiotic cocktail developed a biomass between days 8-14 of embryogenesis and had a reduced hatch rate (t_4=3.572, p=0.023). Hatching rates were unaffected when eggs were treated with the antibiotic cocktail and maintained in a laminar flow hood (t_2=1.289, p=0.326) or under low temperatures (15-20 °C, t_3=0.738, p=0.514) to prevent fungal growth. The antibiotic cocktail included penicillin G, kanamycin, spectinomycin, streptomycin, and gentamicin, each at a concentration of 25 µg/mL. Data point shape reflects eggs taken from the same initial clutch.

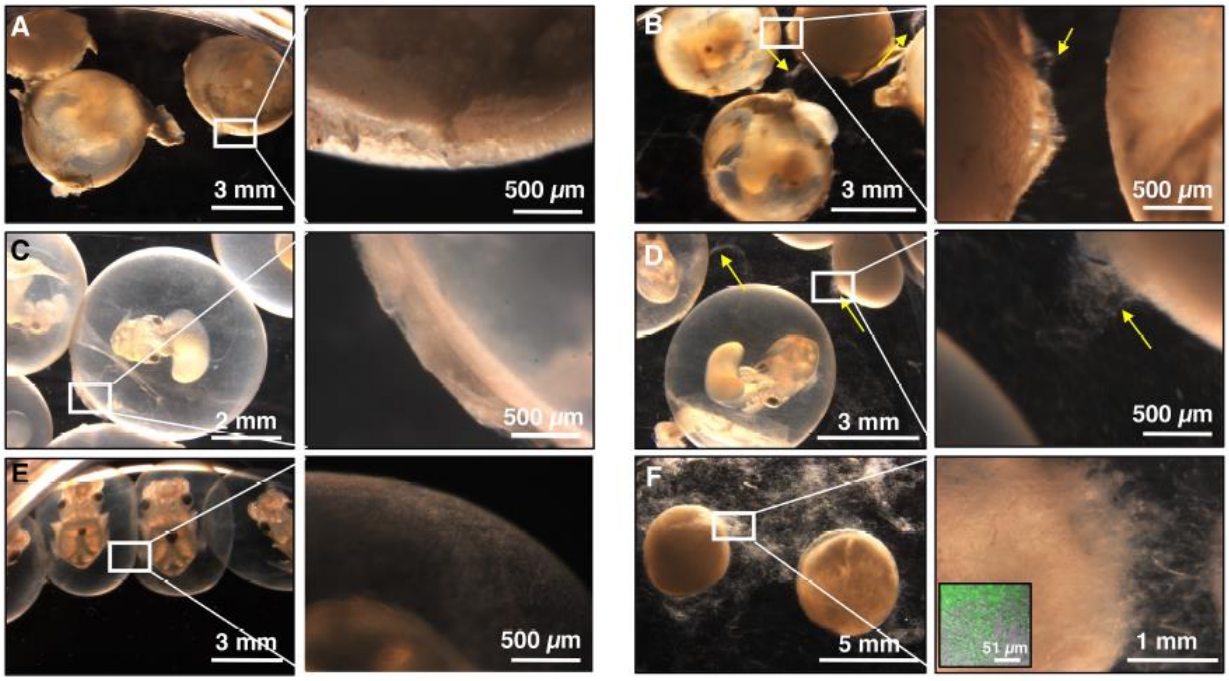


Fig 3. Embryos lacking jelly coats are more susceptible to fungal infection. Eggs were left untreated *(A/C/E)* or were challenged with *F. keratoplasticum* FSSC-2g, 10^4 conidia/mL for 18 days *(B/D/F)*. Eggs were left intact *(A/B)*, or were decapsulated, leaving only the jelly coat and developing embryo *(C/D)* or only the developing embryo *(E/F)*. Hyphae were stained with Syto9 nucleic acid stain and visualized with a confocal microscope (*F*, inset, yellow arrows). (n=4 trials, 10 eggs/treatment).

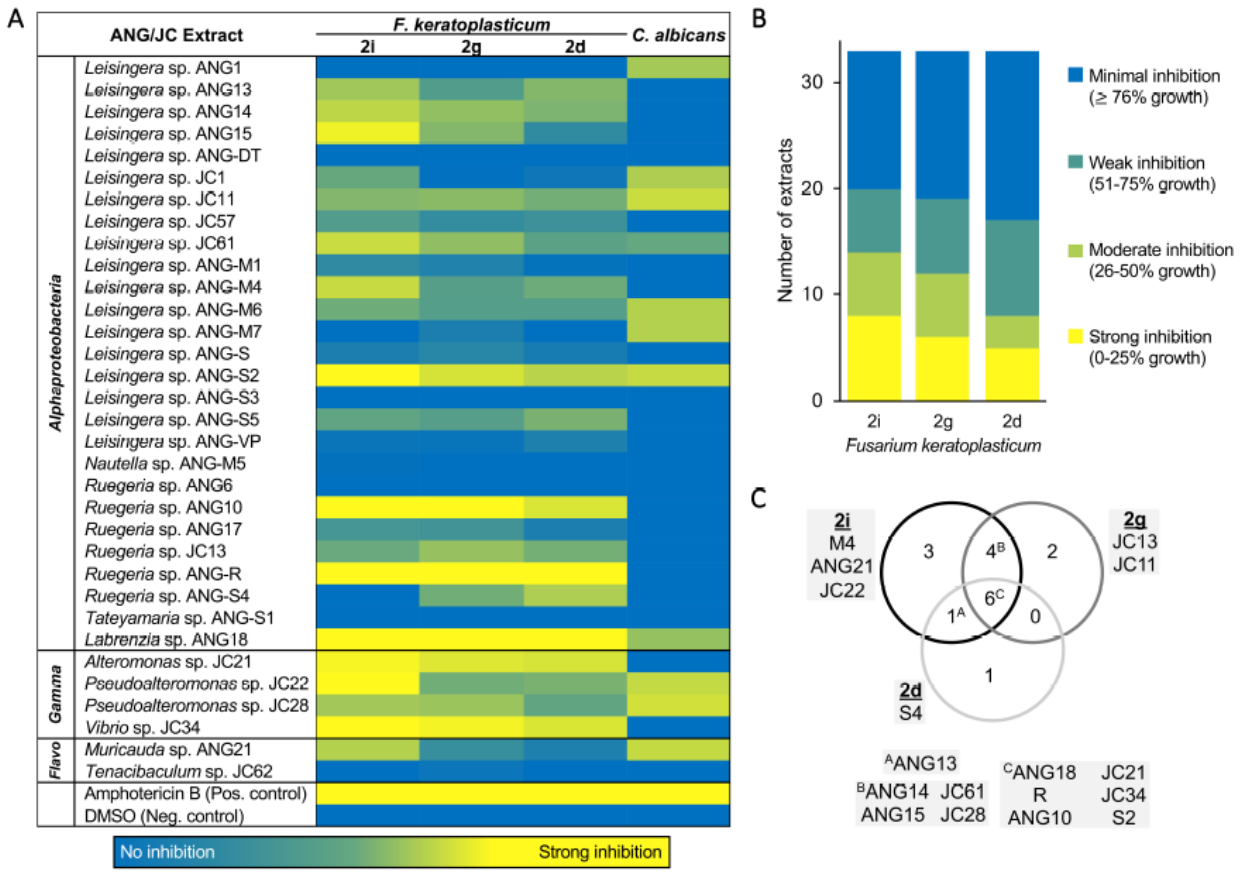


Fig. 4. Bacterial secondary metabolites from the ANG and JC isolates demonstrate differential activity against fungal pathogens. *(A)* Many extracts from both ANG and JC bacterial isolates exhibited antifungal activity against one or more strains of *Fusarium keratoplasticum* and the human pathogen *Candida albicans*. Amphotericin B (4 µg/mL) was used as a positive control. Data shown are based on average % growth values of technical triplicates from at least two experimental replicates. *(B)* Over 51% of ANG and JC bacterial extracts strongly or moderately inhibited the egg-isolated *F. keratoplasticum* strains, FSSC-2i and FSSC-2g compared to the clinical isolate *F. keratoplasticum* FSSC-2d. *(C)* Venn diagram showing extracts that were active against each fusarial strain. Overall, six ANG and JC extracts exhibited antifungal activity against all three *F. keratoplasticum* strains, with other extracts

showing activity against just one or two of the strains. Highlighted boxes indicate strain numbers listed in (*A*).

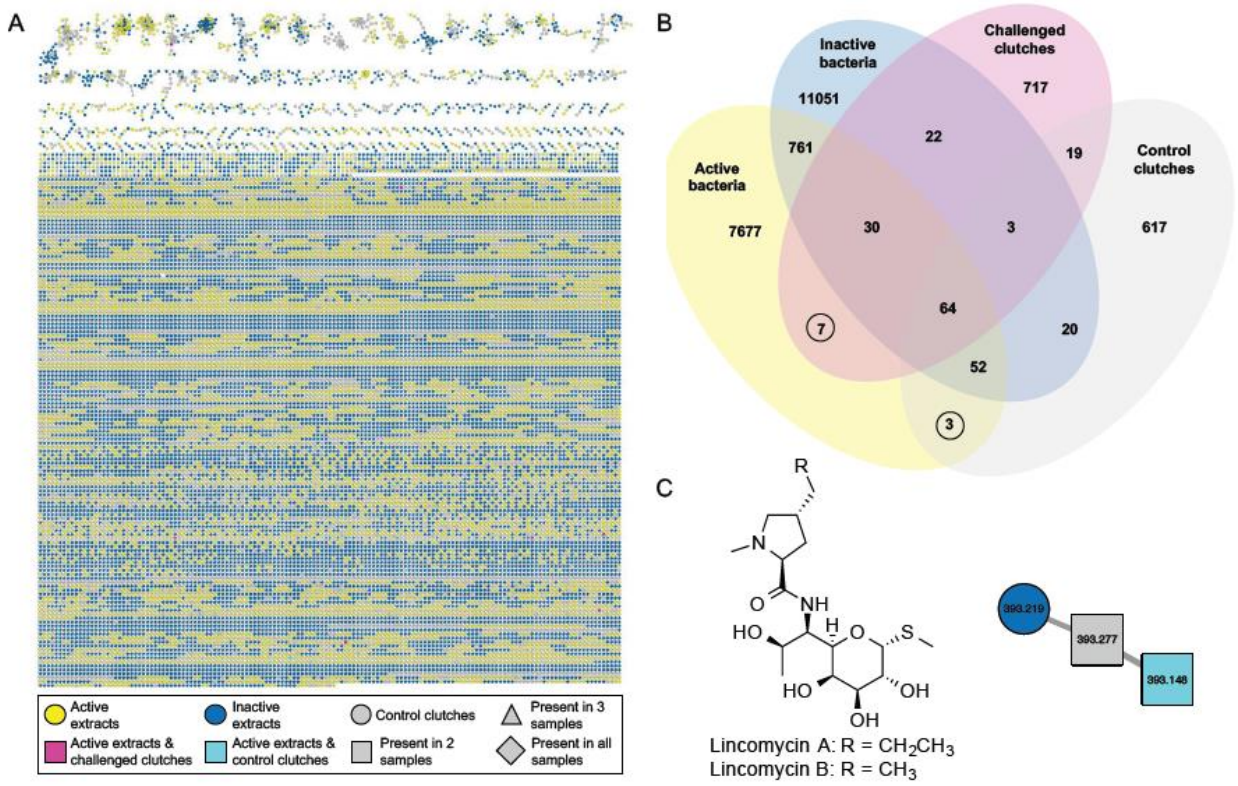


Fig. 5. Comparative metabolomics identifies antimicrobial bacteria-derived secondary metabolites in egg clutches and active ANG and JC bacterial extracts. (*A*) LC-MS/MS molecular network of challenged and control clutches from Fig. 6 with antifusarial active bacterial extracts (yellow circles) and inactive bacterial extracts (blue circles) resulted in a large, complex network (enlarged image of network provided in Fig. S8). Active bacterial extracts include only those with strong inhibition (0-25 % fungal growth, n = 8) while inactive bacterial extracts included only those with minimal to no inhibition (≥ 76 % fungal growth, n = 13). Parent masses are represented within each node and the thickness of the edge is based upon cosine similarity score. (*B*) Ten MS features (circled) were prioritized for further investigation, including seven from challenged clutches that overlap with active bacterial extracts (pink squares) and three from control clutches that overlap with active bacterial extracts (teal squares). (*C*) Lincomycin B ($[M+H]^+$ m/z 393.1480), a structural analogue of the FDA-approved

antimicrobial drug lincomycin A, was identified in both the control clutches and the active extract, *Labrenzia* sp. ANG18 (identification confirmed via comparison with purchased standards; gray squares represent features found in two or more of the remaining categories). Lincomycin A was also found in the control clutch but was not detectable in the *Labrenzia* sp. ANG18 isolate.

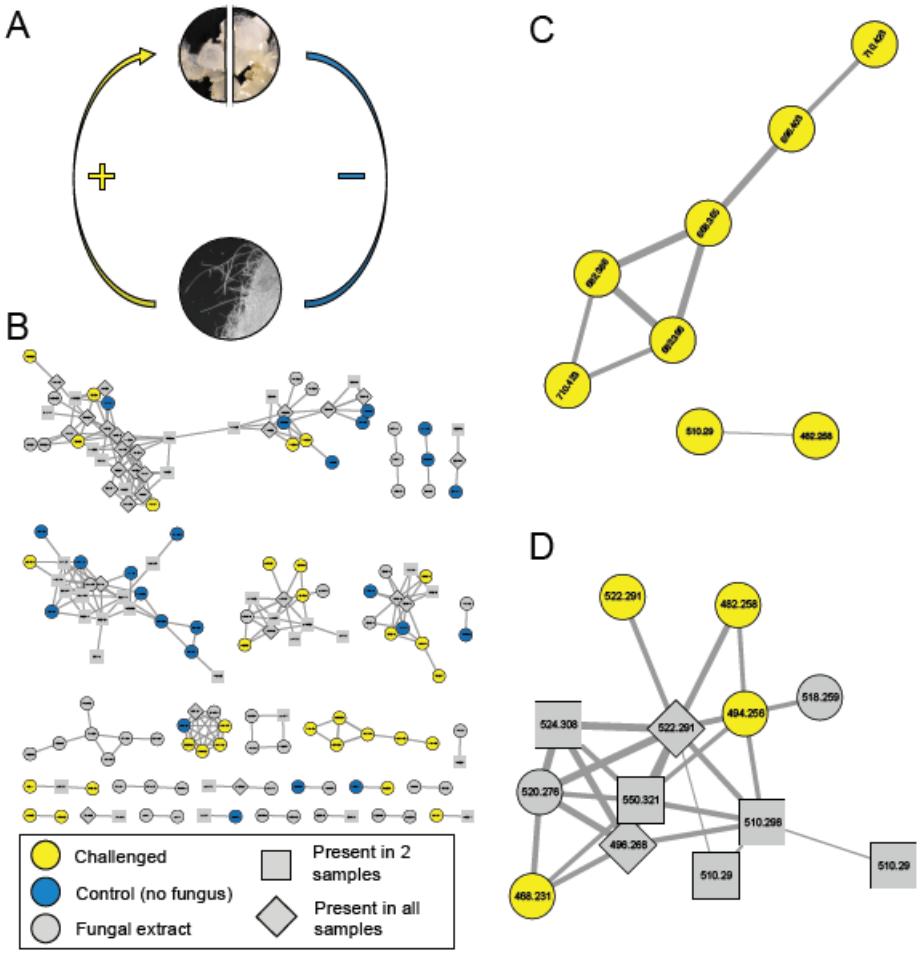


Fig. 6. Comparative metabolomics identified specialized metabolite production by egg clutch-associated bacteria induced in the presence of *Fusarium keratoplasticum*. *(A)* The experimental design involved splitting egg clutches, treating one portion with *F. keratoplasticum* FSSC-2g and leaving one portion as an untreated control. No antibacterial treatment was added, leaving JC-associated bacteria potentially able to metabolically produce defensive metabolites. *(B)* LC-MS/MS molecular networking revealed numerous metabolites present only in challenged clutches (yellow circles), including two clusters with metabolites found only in challenged clutches (*C*). Metabolites were deprioritized if found only in control clutches (blue circles) and/or in *F. keratoplasticum* FSSC-2g extract (gray circles), with features found in two or more

extracts shown in gray squares or diamonds. Parent masses are represented within each node and the thickness of the edge is based upon cosine similarity score. *(D)* Structurally distinct glycerophosphocholines (see text) were also present in another cluster in this network, found in challenged clutches, control clutches, and/or *F. keratoplasticum* FSSC-2g. Several of the metabolites in this cluster are found only in the challenged clutches (yellow circles) or the *F. keratoplasticum* FSSC-2g extract (gray circles), while other metabolites in this cluster are found in both challenged and control clutches (gray squares) or in all three samples (gray diamonds) (Table S4).

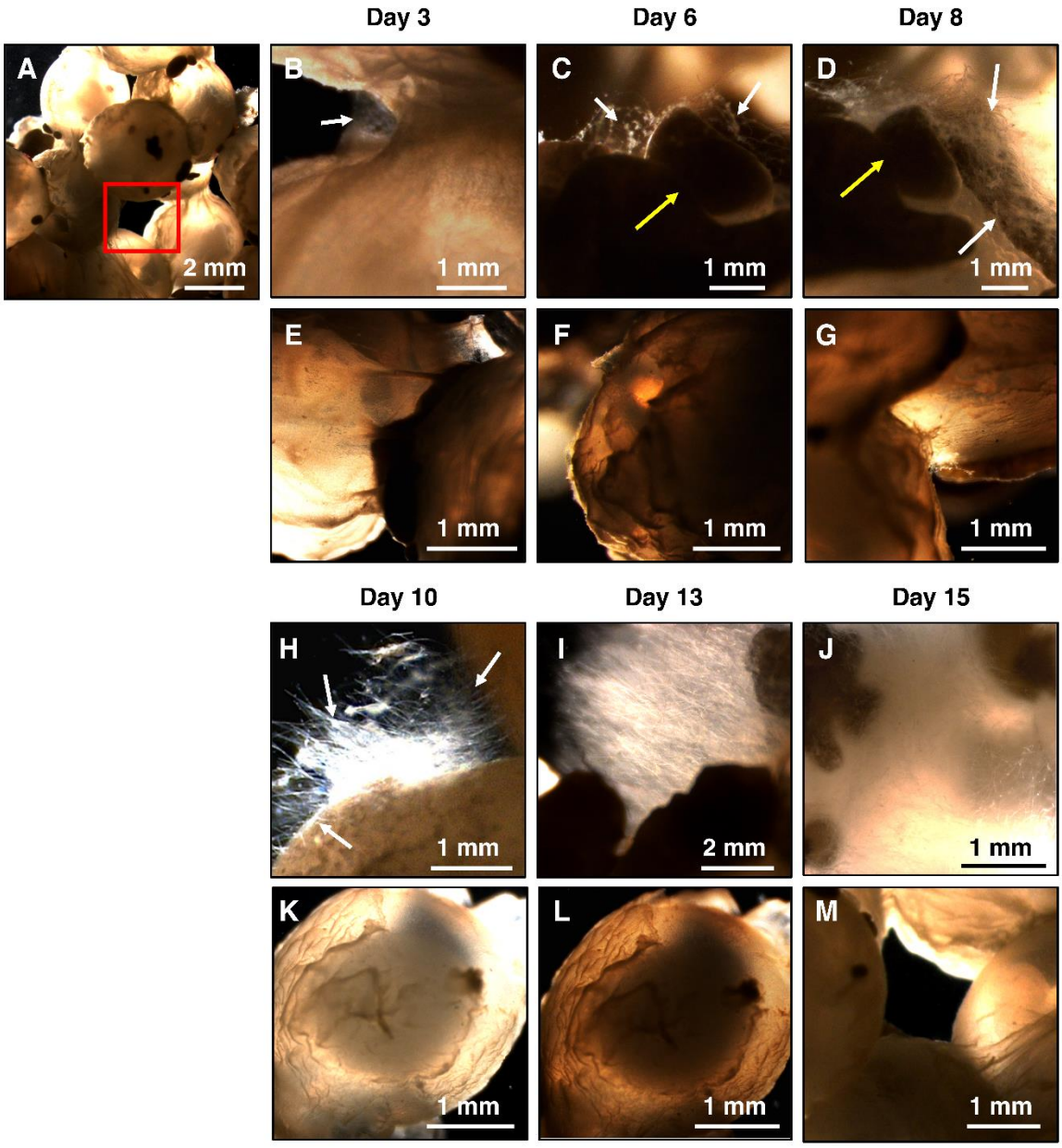
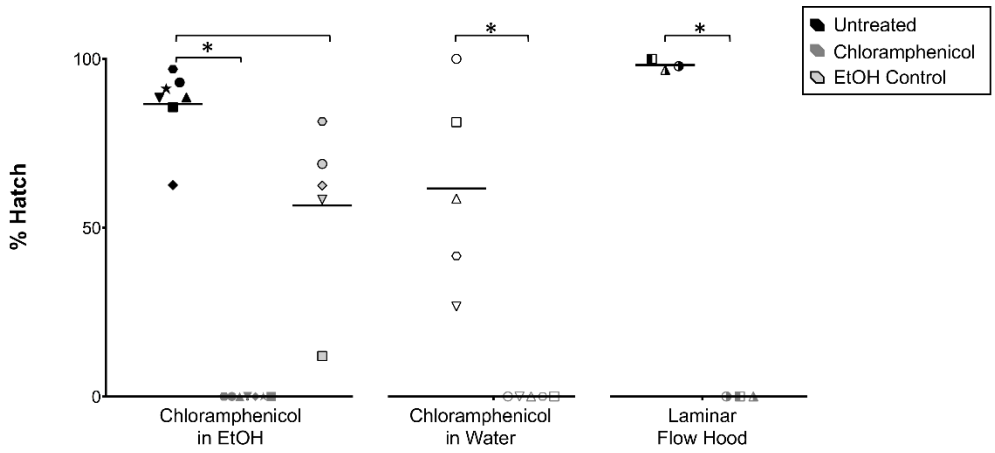


Figure S1. A time series illustrating biofouling development over embryogenesis. Clutches were either treated with antibiotic cocktail (B-D, H-J) or were left untreated (E-G, K-M). Time series images focus on the surface of 1-2 eggs, such as the segment indicated by the red box (A). On day 3, a few hyphae were discernable (B, white arrow), those hyphae increased in number on days 6 (C) and 8 (D), with increased fouling during days 10 (H), 13 (I) and 15 (J). Untreated

eggs showed no signs of fouling or fungal hyphae by day 15 (A, M). The same grain of sand can be seen in C/D, indicated by yellow arrows. Untreated eggs are also shown throughout embryogenesis: day 3 (E), day 6 (F), day 8 (G), day 10 (K), day 13 (L), day 15 (M). Representative images of untreated and antibiotic-treated eggs were taken from the same original clutch.



Experiment Environment	Treatment	# of clutches	Average Eggs/Clutch	Day of Biomass Appearance	Days Hatchlings Appeared	Average % Hatch
Laboratory	Chloramphenicol (20 µg/ml)	7	26	11-19	-	0%
	Ethanol Control	5	29	-	23-27	57% ± 26%
	Untreated	7	38	-	20-27	87% ± 11%
	Chloramphenicol (25 µg/ml)	5	37	10-12	-	0%
	Untreated	5	27	-	17-28	62% ± 30%
Laminar flow hood	Chloramphenicol (25 µg/ml)	3	45	-	-	0%
	Untreated	3	48	-	13 - 23	98% ± 2%

Figure S2. Chloramphenicol treatment of egg clutches resulted in fungal biomass. Eggs that were treated with 20 µg/mL of chloramphenicol dissolved in ethanol (t_6=20.47, p<0.0001), or with 25 µg/mL of chloramphenicol dissolved in FSASW (t_4=4.668, p=0.01), developed fungal fouling in both cases. Treatment with ethanol alone did not significantly reduce juvenile hatching (t_4=2.33, p=0.08). However, hatching levels were affected when eggs were treated with chloramphenicol and maintained in a laminar flow hood (t_2=103.6, p<0.0001) to prevent fungal growth, indicating that chloramphenicol itself interfered with embryo development. Other studies have shown that juvenile squid and adults are not affected by chloramphenicol at similar doses. Data point shape reflects eggs taken from the same initial clutch.

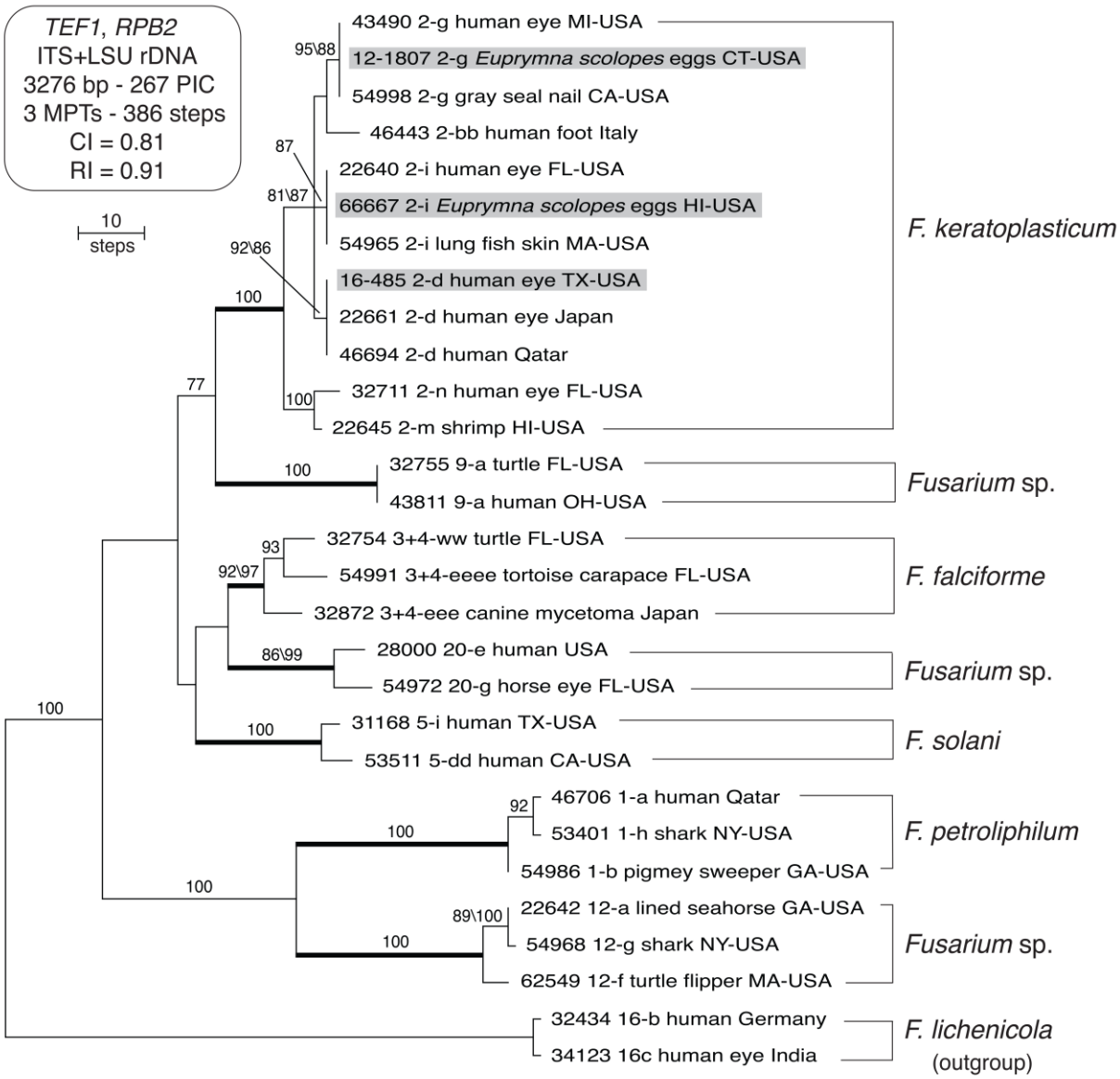


Figure S3. Phylogeny of fungal isolates. One of three most-parsimonious trees (MPTs) inferred from portions of three loci used to identify species and 3-locus haplotypes in the *Fusarium solani* species complex. Arabic numerals and lowercase roman letters, respectively, identify species and unique haplotypes. Numbers above internodes represent the frequency (%) that they were recovered from 1000 maximum likelihood (ML) and maximum parsimony (MP) bootstrap (BS) pseudoreplicates of the data. Only the ML-BS value is shown when it differed by ≤5% of the MP-BS score. Bold internodes are used to identify eight genealogically exclusive ingroup

species, which were rooted on sequences of *F. lichenicola* based on more inclusive analyses (243). The three strains of *F. keratoplasticum* used in the present study are identified by gray highlight. CI, consistency index; PIC, parsimony informative characters; RI, retention index.

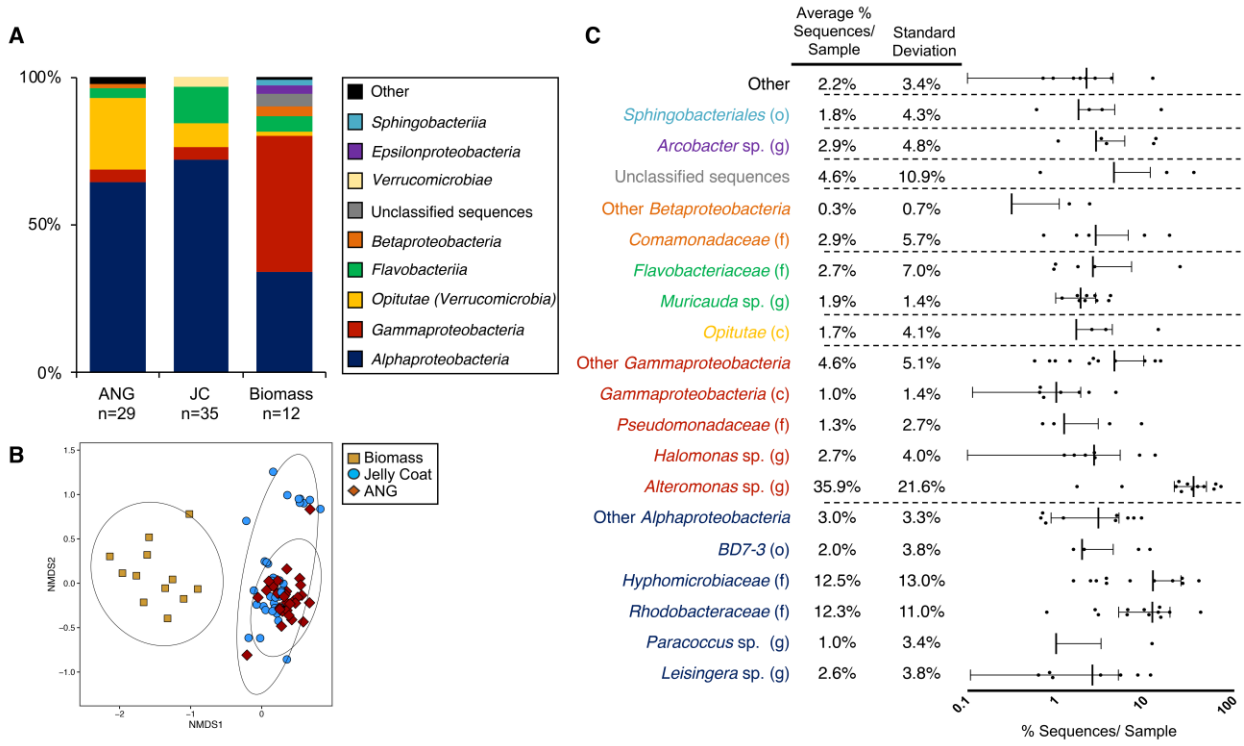


Figure S4. Bacterial diversity in fungal biomass. The fungal biomass bacterial community (n=12) was dominated by *Gammaproteobacteria* (A) and was distinct from that found in the ANG and JC (B). The relative abundances of taxa that make up the fungal biomass community varied substantially between samples (C). Taxa presented at the finest level obtained, c – class; o – order; f – family; g – genus. Mean % sequences/sample represented by thick bar, standard deviation represented by thin bars. Scatter plot is presented on a log scale to demonstrate variation for taxa present at lower average abundances. "Other" includes taxa present at <1% of the average fungal biomass: unclassified *Bacteria*, *Stramenopiles* (o), *Pirellulaceae* (f), unclassified *Proteobacteria* (p), *Myxococcales* (o), and *Nannocystaceae* (f). NMDS plot based on Bray Curtis metric of beta diversity demonstrated that the community composition of the fungal biomass is distinct from that of the ANG/JC (B). ANG and JC data were previously published (72).

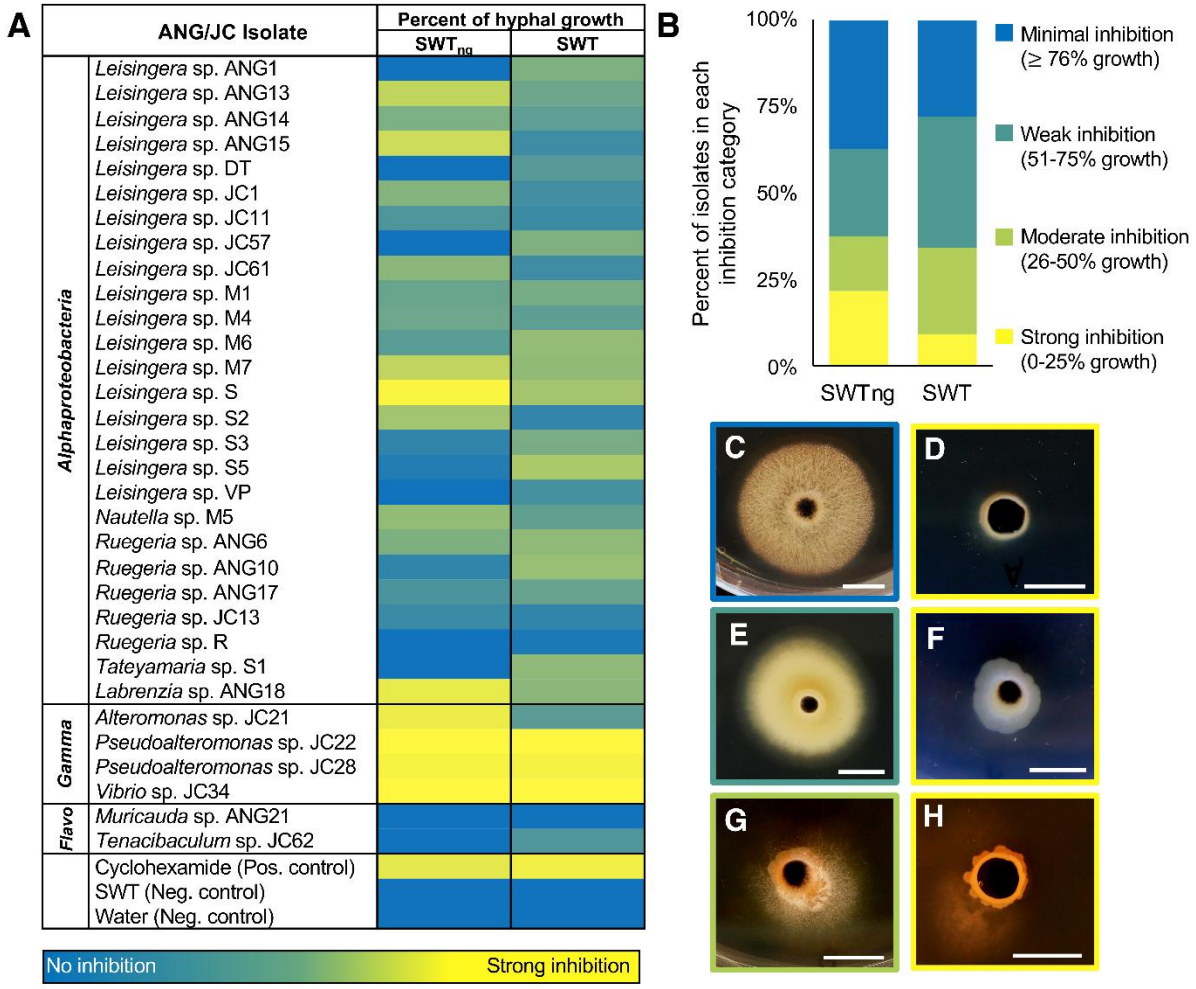


Figure S5. ANG/JC bacterial isolates differentially inhibited *F. keratoplasticum* FSSC-2g. *(A)* Percent of hyphal growth, as compared to the respective media control [seawater tryptone (SWT) or SWT with no glycerol (SWT_{ng})], was measured when conidia were plated in the presence of ANG/JC isolate growth. *(B)* A majority of ANG/JC isolates strongly (0-25 % of control growth), moderately (26-50 % of control), or weakly (51-75 % of control) inhibited fungal growth. Representative images of *F. keratoplasticum* hyphal growth on SWT media *(C)*, in the presence of cycloheximide (1000 µg/mL) on SWT *(D)*, with previous growth from isolate JC21 on SWT *(E)*, JC21 on SWT_{ng} *(F)*, ANG-S5 on SWT *(G)*, and ANG-S5 on SWT_{ng} *(H)*. Scale bars = 1 cm. Average results of three trials. Since bacterial secondary metabolite repression

can occur in the presence of glycerol (252), we tested all of our strains under culture conditions with and without glycerol. Thirteen isolates gained activity (≥ 9 % difference) when grown without glycerol and 13 different isolates lost activity (≥ 9 % difference, Table S2). Although the major carbon sources utilized by bacteria in the JC remain unclear, carbon catabolite repression has been reported in several antimicrobial-producing *Gammaproteobacteria, Firmicutes*, and *Actinobacteria* (252). The effects of carbon source on secondary metabolite production are pathway specific, however, and the presence of glycerol has been shown to increase production of the antifungal nystatin in *Streptomyces noursei* (282). Because little is known about the chemistry of marine *Proteobacteria* and their secondary metabolite biosynthesis, various culture conditions should be tested in future studies to better explore the biosynthetic potential of ANG/JC bacteria.

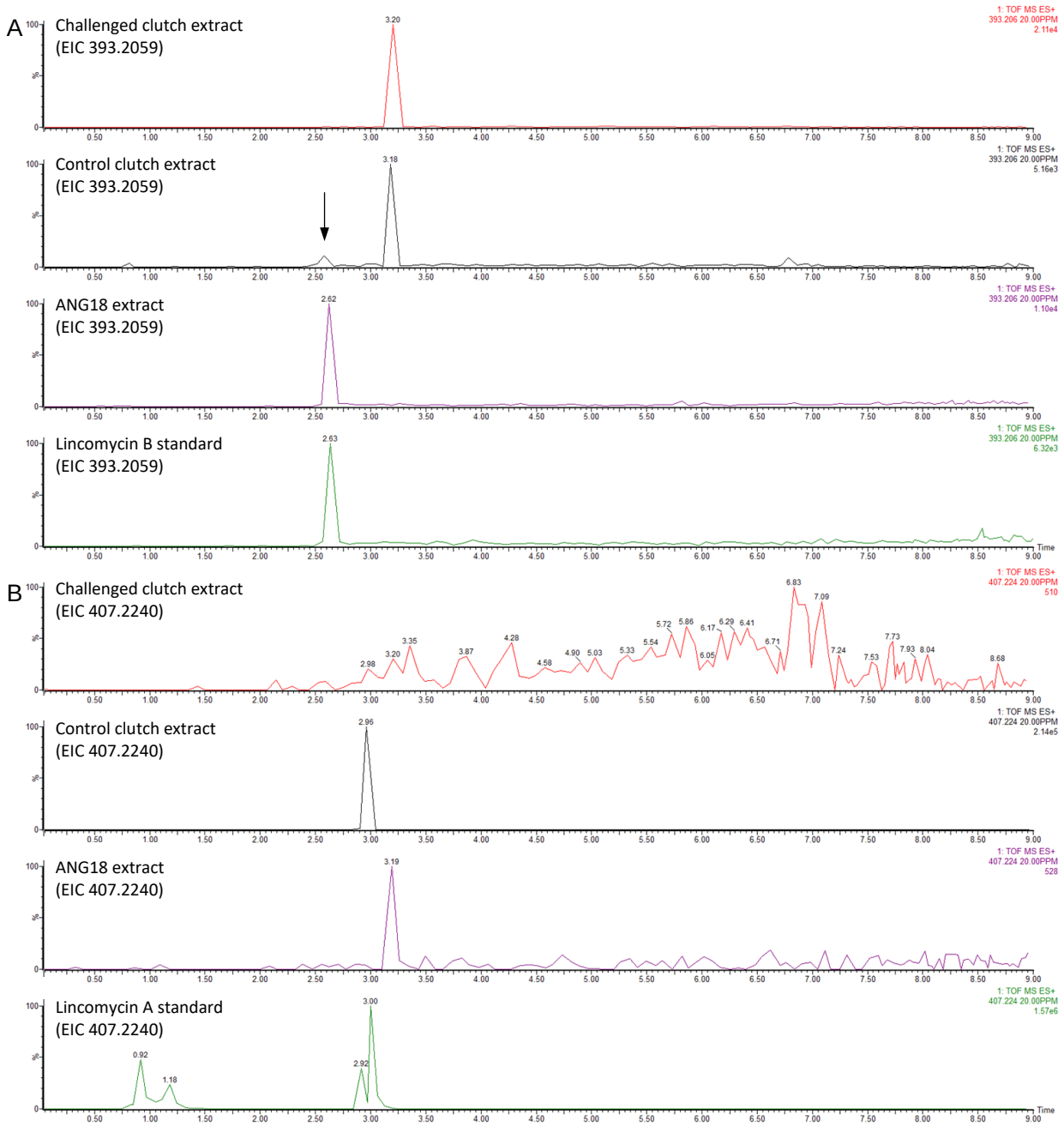
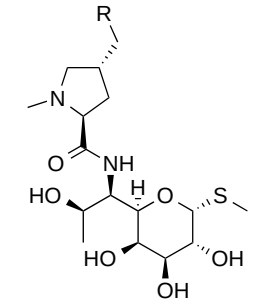
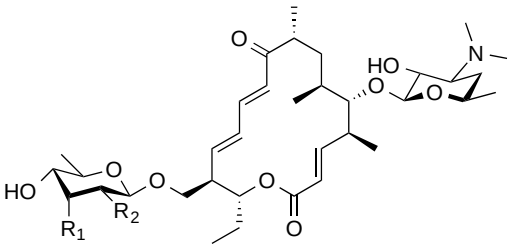


Figure S6. Lincomycin LC-MS/MS analyses. Confirmation of the presence of lincomycins A and B (arrow) in the control clutch extract including extracted ion chromatograms (EICs) for lincomycin B (Fig. S6A; $[M+H]^+$ *m/z* 393.2059, t_R 2.63 min) and lincomycin A (Fig. S6B; $[M+H]^+$ *m/z* 407.2240, t_R 2.95 min; the lincomycin A standard splits into two peaks but the t_R and fragmentation patterns match for the standard and the matching peak in the control clutch; the

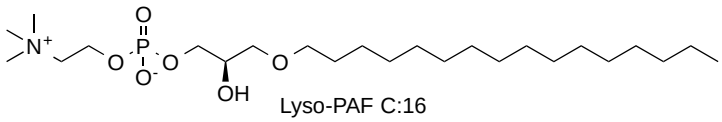
peak at t_R 3.2 min was not consistent with lincomycin fragmentation). Lincomycin B was also detected in the bacterial isolate *Labrenzia* sp. ANG18 although lincomycin A was not detected in this isolate. Lincomycin A and B standards (Santa Cruz Biotechnology, Inc., Dallas, TX) were authenticated using 1H NMR and HRMS prior to LC-MS/MS comparison.



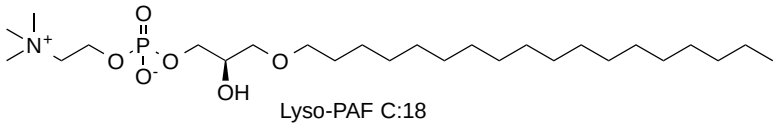
Lincomycin A: R = CH_2CH_3
Lincomycin B: R = CH_3



Mycinamicin III: R_1 = OH, R_2 = OMe
Mycinamicin IV: R_1 = OMe, R_2 = OMe
Mycinamicin VI: R_1 = OH, R_2 = OH



Lyso-PAF C:16



Lyso-PAF C:18

Compound/ Cluster	**Sample(s)**	**Node mass[a]**	**Observed *m/z* $[M+H]^+$ [b]**	**Theoretical *m/z* $[M+H]^+$**
Lincomycin A	Control clutches	407.146	407.2267	407.2216
Lincomycin B	Control clutches and *Leisingera* sp. ANG18	393.151	393.2054	393.2059
Fig. 6C, Cluster 1	Challenged clutches	668.365	668.4087	
Fig. 6C, Cluster 1[c]	Challenged clutches	682.386	682.4274	
Fig. 6C, Cluster 1	Challenged clutches	696.403	696.4419	
Fig. 6C, Cluster 1 [c]	Challenged clutches	710.423	710.4592	
Fig. 6C, Cluster 2	Challenged clutches	482.258	482.3294	
Fig. 6C, Cluster 2	Challenged clutches	510.290	510.3581	
Mycinamicin III				682.4166
Mycinamicin IV				696.4323
Mycinamicin VI				668.4010
Lyso-PAF C:16				482.3610
Lyso-PAF C:18				510.3923

[a] consensus mass generated through GNPS

[b] experimental spectra obtained using Waters Synapt G2-Si, processed with MassLynx V4.1

[c] two nodes with identical mass are found in this cluster, likely representing isomers

Figure S7. Structures and sample sources of lincomycins, mycinamicins, and lyso-PAFs identified and/or tested herein.

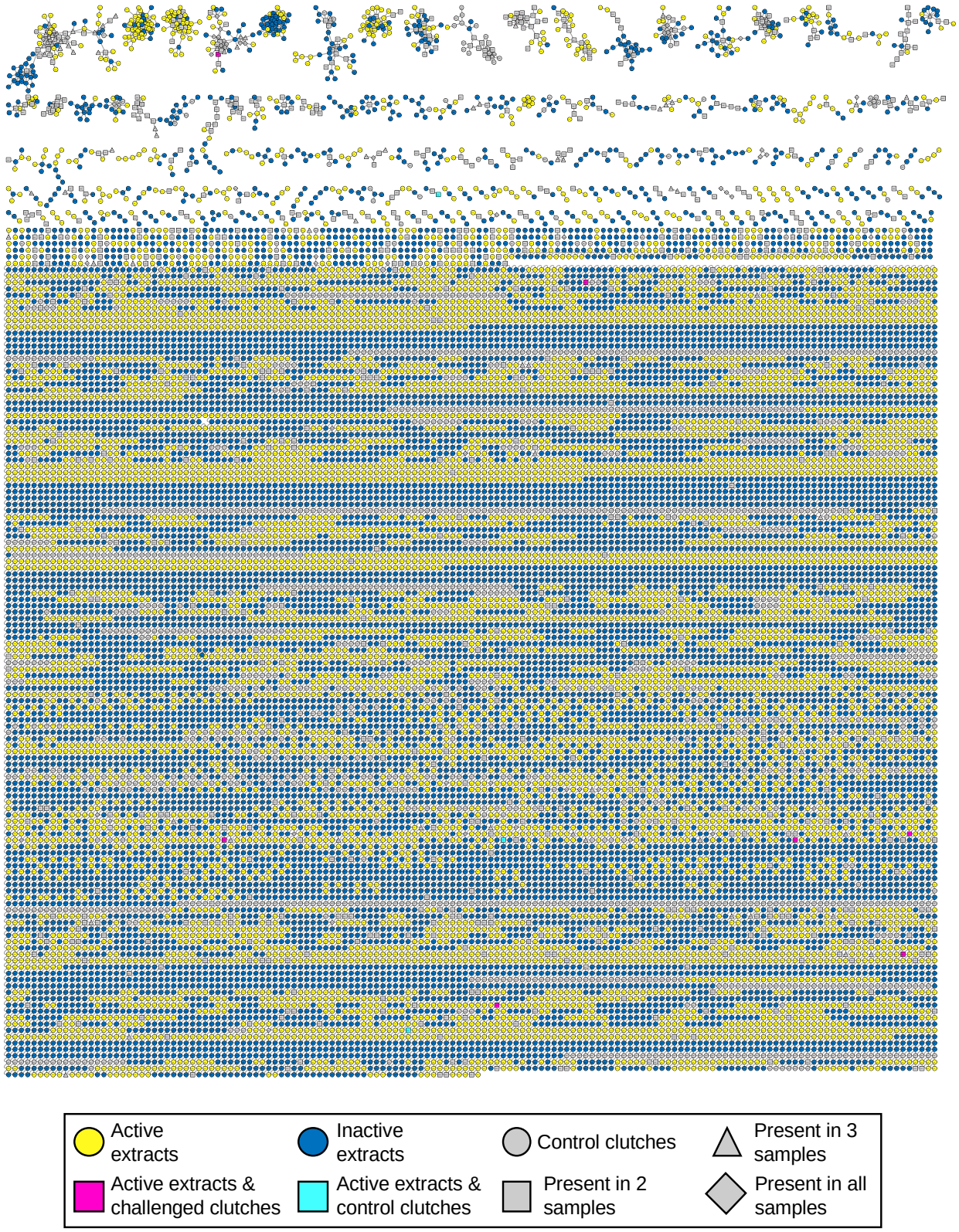


Figure S8. Enlarged image of LC-MS/MS molecular network of challenged and control clutches and active and inactive bacteria from Fig. 5A. LC-MS/MS molecular network of

challenged and control clutches with antifusarial active bacterial extracts (yellow circles) and inactive bacterial extracts (blue circles) resulted in a large, complex network. Active bacterial extracts include only those with strong inhibition (0-25 % fungal growth, n = 8) while inactive bacterial extracts included only those with minimal to no inhibition (≥ 76 % fungal growth, n = 13). Parent masses are represented within each node and the thickness of the edge is based upon cosine similarity score.

Table S1. Bacterial and fungal strains used in this study.

Strain	Characteristics	Source
Fusarium keratoplasticum FSSC-2g	CT isolated strain	This study
Fusarium keratoplasticum FSSC-2i	HI isolated strain	This study
Fusarium keratoplasticum FSSC-2d	Texas isolated strain	Clinical isolate; this study
Candida albicans ATCC 18804	Type strain	American Type Culture Collection (ATCC)
Leisingera sp. ANG1	Isolate from *E. scolopes* ANG	Collins and Nyholm 2011, Collins et al., 2015
Leisingera sp. ANG13	Isolate from *E. scolopes* ANG	This study
Leisingera sp. ANG14	Isolate from *E. scolopes* ANG	This study
Leisingera sp. ANG15	Isolate from *E. scolopes* ANG	This study
Leisingera sp. ANG-DT	Isolate from *E. scolopes* ANG	Collins et al. 2015
Leisingera sp. ANG-M1	Isolate from *E. scolopes* ANG	Collins et al. 2015
Leisingera sp. ANG-M4	Isolate from *E. scolopes* ANG	Collins et al. 2012
Leisingera sp. ANG-M6	Isolate from *E. scolopes* ANG	Collins et al. 2015
Leisingera sp. ANG-M7	Isolate from *E. scolopes* ANG	Collins et al. 2015
Leisingera sp. ANG-S	Isolate from *E. scolopes* ANG	Collins et al. 2015
Leisingera sp. ANG-S2	Isolate from *E. scolopes* ANG	Collins et al. 2012
Leisingera sp. ANG-S3	Isolate from *E. scolopes* ANG	Collins et al. 2015
Ruegeria sp. ANG-S4	Isolate from *E. scolopes* ANG	Collins et al. 2015
Leisingera sp. ANG-S5	Isolate from *E. scolopes* ANG	Collins et al. 2015
Leisingera sp. ANG-VP	Isolate from *E. scolopes* ANG	Collins et al. 2015
Leisingera sp. JC1	Isolate from *E. scolopes* JC	Gromek et al. 2016
Leisingera sp. JC11	Isolate from *E. scolopes* JC	This study
Leisingera sp. JC57	Isolate from *E. scolopes* JC	This study
Leisingera sp. JC61	Isolate from *E. scolopes* JC	This study
Labrenzia sp. ANG18	Isolate from *E. scolopes* ANG	This study
Nautella sp. ANG-M5	Isolate from *E. scolopes* ANG	Collins et al. 2012
Ruegeria sp. ANG6	Isolate from *E. scolopes* ANG	This study
Ruegeria sp. ANG10	Isolate from *E. scolopes* ANG	This study
Ruegeria sp. ANG17	Isolate from *E. scolopes* ANG	This study
Ruegeria sp. ANG-R	Isolate from *E. scolopes* ANG	Collins et al. 2015
Ruegeria sp. JC13	Isolate from *E. scolopes* ANG	This study
Tateyamaria sp. ANG-S1	Isolate from *E. scolopes* ANG	Collins et al. 2015
Muricauda sp. ANG21	Isolate from *E. scolopes* ANG	Gromek et al. 2016
Tenacibaculum sp. JC62	Isolate from *E. scolopes* JC	This study
Altermonas sp. JC21	Isolate from *E. scolopes* JC	This study
Pseudoaltermonas sp. JC22	Isolate from *E. scolopes* JC	This study
Pseudoaltermonas sp. JC28	Isolate from *E. scolopes* JC	This study
Vibrio sp. JC34	Isolate from *E. scolopes* JC	This study

Table S2. Percentages of *F. keratoplasticum* FSSC-2g hyphal growth from well diffusion assay.

	ANG/JC Isolate	Percent Hyphal Growth	
		SWT_{ng}	**SWT**
Alphaproteobacteria	*Leisingera* sp. ANG1	111.1±0.42	54±1.2
	Leisingera sp. ANG13	25.7±0.11	61.7±1.45
	Leisingera sp. ANG14	54.2±0.24	68±0.24
	Leisingera sp. ANG15	21±0.66	81±1.2
	Leisingera sp. ANG-DT	117.3±0.33	70.8±0.77
	Leisingera sp. JC1	51.4±0.45	79.2±1.1
	Leisingera sp. JC11	74.8±0.32	82.1±0.3
	Leisingera sp. JC57	116.4±2.33	53.9±0.06
	Leisingera sp. JC61	48.8±0.15	81±1.26
	Leisingera sp. ANG-M1	63.2±0.86	56.8±2.8
	Leisingera sp. ANG-M4	60.7±0.09	67.3±0.76
	Leisingera sp. ANG-M6	69.7±1.34	45.2±1.5
	Leisingera sp. ANG-M7	26.2±0.46	46±0.26
	Leisingera sp. ANG-S	2.3±0.23	38±1.1
	Leisingera sp. ANG-S2	39±0.28	86±0.62
	Leisingera sp. ANG-S3	86.7±0.13	56.1±0.15
	Leisingera sp. ANG-S5	92.9±0.27	35.3±0.7
	Leisingera sp. ANG-VP	125±0.96	77.7±1.2
	Nautella sp. ANG-M5	45.5±1.35	66.6±0.34
	Ruegeria sp. ANG6	54±0.24	46±1.7
	Ruegeria sp. ANG10	87.1±0.43	43±1.5
	Ruegeria sp. ANG17	75.4±0.2	64.3±0.96
	Ruegeria sp. JC13	82.6±0.46	86.7±0.81
	Ruegeria sp. ANG-R	105.9±1.5	95.6±0.26
	Tateyamaria sp. ANG-S1	112.1±1.5	47±1.2
	Labrenzia sp. ANG18	9.42±0.22	49.1±0.29
Gamma.	*Alteromonas* sp. JC21	7.6±0.23	69.7±0.44
	Pseudoalteromonas sp. JC22	0±0	0±0
	Pseudoalteromonas sp. JC28	3.1±0.1	4.4±0.11
	Vibrio sp. JC34	0±0	0±0
Flavo.	*Muricauda* sp. ANG21	155.1±0.63	110.8±0.42
	Tenacibaculum sp. JC62	132.7±0.31	73.9±1.4
Controls	Cyclohexamide (Pos. control)	11.2±0.1	5.1±0.21
	SWT (Neg. control)	100±0.68	100±0.46
	Water (Neg. control)	109.8±0.95	100.4±0.36

Percentages are an average of at least 3 trials ± the standard error of the mean. *Gamma.* (*Gammaproteobacteria*),*Flavo.*(*Flavobacteriia*).

Table S3. Percentages of *F. keratoplasticum* spp. (FSSC-2i, FSSC-2g, and FSSC-2d) and *Candida albicans* wild type (Ca) fungal growth from the 96-well liquid antifungal assay.

	ANG/JC Isolate	Percent Fungal Growth			
		FSSC-2i	FSSC-2g	FSSC-2d	Ca
Alphaproteobacteria	*Leisingera* sp. ANG1	104.3±5.3	106.8±14.9	116.3±14.5	36.7±5.5
	Leisingera sp. ANG13	42.8±1.1	68.0±1.6	49.4±2.5	141.2±10.2
	Leisingera sp. ANG14	33.1±8.2	42.8±8.7	52.2±2.8	132.9±1.5
	Leisingera sp. ANG15	14.8±14.8	48.4±3.2	81.2±4.8	158.9±6.1
	Leisingera sp. ANG-DT	108.3±8.3	109.3±5.6	134.0±4.3	136.7±1.3
	Leisingera sp. JC1	64.0±9.5	99.6±7.5	94.6±11.1	33.3±9.8
	Leisingera sp. JC11	53.5±5.9	45.9±4.8	56.4±2.5	22.5±3.9
	Leisingera sp. JC57	71.4±6.8	79.5±11.4	76.0±5.3	116.3±8.9
	Leisingera sp. JC61	29.1±2.5	43.6±8.1	64.9±14.5	61.2±18.9
	Leisingera sp. ANG-M1	83.3±3.2	87.4±4.8	97.9±13.3	153.4±3.0
	Leisingera sp. ANG-M4	29.8±3.1	66.3±7.7	58.6±0.6	142.0±15.0
	Leisingera sp. ANG-M6	61.0±6.5	66.8±6.4	67.5±10.7	31.2±7.8
	Leisingera sp. ANG-M7	101.2±8.4	89.7±9.6	121.8±10.9	31.4±7.4
	Leisingera sp. ANG-S	90.7±7.1	84.3±5.6	91.5±6.2	152.0±5.1
	Leisingera sp. ANG-S2	-11.7±9.3	16.8±15.0	29.6±15.4	24.2±7.0
	Leisingera sp. ANG-S3	118.7±11.6	138.5±5.5	122.0±11.7	183.8±0.2
	Leisingera sp. ANG-S5	65.8±13.7	66.4±8.8	53.4±8.0	112.0±2.7
	Leisingera sp. ANG-VP	96.5±5.7	97.2±1.5	88.5±12.1	144.7±9.9
	Nautella sp. ANG-M5	98.8±12.3	106.5±4.1	115.5±1.2	165.1±10.7
	Ruegeria sp. ANG6	104.7±5.4	117.8±10.1	104.9±8.9	137.3±12.2
	Ruegeria sp. ANG10	8.6±13.7	-15.1±29.0	16.6±13.2	122.2±3.6
	Ruegeria sp. ANG17	75.5±0.5	74.6±6.5	89.9±0.1	102.7±1.8
	Ruegeria sp. JC13	62.9±5.6	41.1±12.1	55.3±10.4	146.3±10.2
	Ruegeria sp. ANG-R	-150.9±22.7	-230.0±3.7	-109.6±14.7	110.0±10.5
	Ruegeria sp. ANG-S4	118.2±2.6	55.9±3.4	34.1±1.4	168.7±3.3
	Tateyamaria sp. ANG-S1	134.3±8.3	125.1±8.8	134.7±14.7	163.9±0.2
	Labrenzia sp. ANG18	-23.3±13.2	-60.9±18.8	-18.9±13.9	42.4±14.3
Gamma.	*Alteromonas* sp. JC21	12.7±14.8	13.7±21.2	18.4±8.8	164.3±5.3
	Pseudoalteromonas sp. JC22	-89.6±3.2	55.5±8.0	53.3±14.6	24.7±7.4
	Pseudoalteromonas sp. JC28	42.3±0.5	39.3±8.1	62.6±9.5	19.8±4.2
	Vibrio sp. JC34	7.8±10.8	4.9±11.1	16.4±7.6	116.7±3.7
Flavo.	*Muricauda* sp. ANG21	36.5±11.7	78.8±5.9	89.0±8.4	24.3±3.0
	Tenacibaculum sp. JC62	116.3±0.7	96.0±6.4	123.9±1.4	127.0±6.0
Controls	Amp B (Pos. control)	9.7±4.2	0.3±0.3	2.5±1.4	2.0±0.4
	DMSO (Neg. control)	100±0.0	100±0.0	100±0.0	100±0.0

Percentages are an average of at least 2 experimental replicates performed in technical triplicate ± the standard error of the mean. *Gamma*. (*Gammaproteobacteria*), *Flavo.* (*Flavobacteriia*).

Table S4. Cluster of additional phosphocholines identified through the GNPS library from challenged clutches, control clutches, and *Fusarium keratoplasticum* FSSC-2g extracts.

Putative Identification[a]	Sample	Node mass[b]	Observed *m/z* $[M+H]^+$ [c]	Theoretical *m/z* $[M+H]^+$	Mass accuracy (ppm)
PC 0:0/14:0	Challenged clutches	468.231	468.3049	468.3085	7.6
PC 15:0/0:0	Challenged clutches	482.258	482.3198	482.3241	8.9
PC 16:1/0:0	Challenged clutches	494.256	494.3265	494.3241	4.8
PC 0:0/18:1[d]	Challenged clutches	522.291	522.3575	522.3554	4.0
PC 18:2/0:0	*F. keratoplasticum* FSSC-2g	520.276	520.3347	520.3398	9.7
PC 18:3/0:0	*F. keratoplasticum* FSSC-2g	518.259	518.3257	518.3252	0.9
PC 17:0/0:0[d]	Challenged and control clutches	510.290	510.3581	510.3554	5.3
PC 17:0/0:0[d]	Challenged and control clutches	510.290	510.3581	510.3554	5.3
PC 20:1/0:0	Challenged and control clutches	550.321	550.3832	550.3867	6.4
PC 17:0/0:0[d]	Challenged and control clutches	510.298	510.3581	510.3554	5.3
PC 0:0/18:0	Challenged and control clutches	524.308	524.3744	524.3711	6.4
PC 0:0/16:0	All 3	496.268	496.3372	496.3398	5.2
PC 0:0/18:1[d]	All 3	522.291	522.3575	522.3554	4.0

[a] features could represent one of several related isomers

[b] consensus mass generated through GNPS

[c] experimental spectra obtained using Waters Synapt G2-Si, processed with MassLynx V4.1

[d] several nodes with similar masses were found in this cluster, likely representing isomers

Table S5. Mass spectral ion intensities for nodes/compounds of interest from challenged and control clutches.

	Node/Compound							
	Lyso-PAF-like				Mycinamicin-like			
Sample	**482.258**	**510.290**	**C:16[a]**	**C:18[a]**	**668.365**	**682.386**	**696.403**	**710.423**
Control clutch 1[b]	2.62e3[c]	4.60e3	9.25e4	1.77e4	3.59e3[c]	2.39e3[c]	1.61e3[c]	2.08e3[c]
Challenged clutch 1[b]	3.45e3	7.58e3	9.23e4	6.60e4	1.24e4	6.43e3	1.20e4	1.03e4
Fold change (challenged)	1.3	1.6	1.0	3.7	3.5	2.7	7.5	5.0
Control clutch 2	3.50e4	4.47e4	6.27e5	3.45e5	9.76e3	6.12e3	1.33e3	1.65e3
Challenged clutch 2	2.25e4	5.29e4	3.55e5	3.46e5	2.76e4	8.75e4	6.58e3	7.17e3
Fold change (challenged)	0.6	1.2	0.6	1.0	2.8	14.3	4.9	4.3
Control clutch 3	4.50e4	2.55e4	5.33e5	3.55e5	4.24e4	3.44e3	9.05e3	7.21e3
Challenged clutch 3	8.50e4	3.60e4	2.76e5	3.39e5	1.52e4	1.48e4	2.19e3	4.06e3
Fold change (challenged)	1.9	1.4	0.5	1.0	0.4	4.3	0.2	0.6
Control clutch 4	3.13e4	1.73e4	2.79e5	5.50e4	3.06e4	1.28e4	4.79e3	3.89e3
Challenged clutch 4	5.51e4	7.49e3	1.95e5	2.75e5	1.41e4	7.46e3	4.41e3	2.16e3
Fold change (challenged)	1.8	0.4	0.7	5.0	0.5	0.6	0.9	0.6
Mean fold change (challenged; ± SEM)	1.4±0.3	1.2±0.3	0.7±0.1	2.7±1.0	1.8±0.8	5.5±3.0	3.4±1.7	2.6±1.2

[a] Lyso-PAF C:16 and lyso-PAF C:18

[b] Clutch 1 was used for GNPS molecular networking in Fig. 6

[c] Below limit for MS/MS and thus not detected in GNPS although found to be present via MassLynx during MS scan

Appendix II

Symbiotic organs shaped by distinct modes of genome evolution in cephalopods*

Significant contributions:

This work was a hugely collaborative effort to sequence the genome and assemble one comprehensive transcriptome of the Hawaiian bobtail squid, *Euprymna scolopes*. My contribution was the sequencing of host RNA from ANG tissue, assisting with analysis of ANG transcripts featured in Figure 3, and assisted with editing the manuscript.

***Citation:**

Madhi Baelcaid, Giorgio Casaburi, Sarah J. McAnulty, Hannah Schmidbaur, Andrea M. Suria, Silvia Moriano-Gutierrez, M. Sabrina Pankey, Todd H.Oakley, Natacha Kremer, Eric J. Koch, Andrew J. Collins, Hoan Nguyen, Sai Lek, Irina Goncharenko-Foster, Patrick Minx, Erica Sodergren, GeorgeWeinstock, Daniel S. Rokhsar, Margaret McFall-Ngai, Oleg Simakov, Jamie S. Foster, Spencer V. Nyholm. (2019) Symbiotic organs shaped by distinct modes of genome evolution in cephalopods *Proceedings of the National Academy of Sciences*. 116 (8): 3030-3035; DOI:10.1073/pnas.1817322116

Symbiotic organs shaped by distinct modes of genome evolution in cephalopods

Mahdi Belcaid[a], Giorgio Casaburi[b], Sarah J. McAnulty[c], Hannah Schmidbaur[d], Andrea M. Suria[c], Silvia Moriano-Gutierrez[a], M. Sabrina Pankey[e], Todd H. Oakley[e], Natacha Kremer[f], Eric J. Koch[a], Andrew J. Collins[c], Hoan Nguyen[g], Sai Lek[g], Irina Goncharenko-Foster[b], Patrick Minx[h], Erica Sodergren[g], George Weinstock[g], Daniel S. Rokhsar[i,j,k], Margaret McFall-Ngai[a], Oleg Simakov[d,l,1], Jamie S. Foster[b,1], and Spencer V. Nyholm[c,1]

[a]Pacific Biosciences Research Center, University of Hawaii, Honolulu, HI 96822; [b]Department of Microbiology and Cell Science, Space Life Science Lab, University of Florida, Merritt Island, FL 32953; [c]Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269; [d]Department of Molecular Evolution and Development, University of Vienna, 1090 Vienna, Austria; [e]Ecology, Evolution, and Marine Biology Department, University of California, Santa Barbara, CA 93106; [f]Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, Claude Bernard University Lyon 1, CNRS, UMR 5558, 69100 Villeurbanne, France; [g]Jackson Laboratory for Genomic Medicine, Farmington, CT 06032; [h]McDonnell Genome Institute, Washington University, St. Louis, MO 63108; [i]Molecular Genetics Unit, Okinawa Institute of Science and Technology, Okinawa 904-0495, Japan; [j]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; and [k]Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598

Edited by Nancy A. Moran, University of Texas at Austin, Austin, TX, and approved December 11, 2018 (received for review October 9, 2018)

Microbes have been critical drivers of evolutionary innovation in animals. To understand the processes that influence the origin of specialized symbiotic organs, we report the sequencing and analysis of the genome of *Euprymna scolopes*, a model cephalopod with richly characterized host–microbe interactions. We identified large-scale genomic reorganization shared between *E. scolopes* and *Octopus bimaculoides* and posit that this reorganization has contributed to the evolution of cephalopod complexity. To reveal genomic signatures of host–symbiont interactions, we focused on two specialized organs of *E. scolopes*: the light organ, which harbors a monoculture of *Vibrio fischeri*, and the accessory nidamental gland (ANG), a reproductive organ containing a bacterial consortium. Our findings suggest that the two symbiotic organs within *E. scolopes* originated by different evolutionary mechanisms. Transcripts expressed in these microbe-associated tissues displayed their own unique signatures in both coding sequences and the surrounding regulatory regions. Compared with other tissues, the light organ showed an abundance of genes associated with immunity and mediating light, whereas the ANG was enriched in orphan genes known only from *E. scolopes*. Together, these analyses provide evidence for different patterns of genomic evolution of symbiotic organs within a single host.

cephalopods | symbiosis | evolution | genomics | transcriptomics

Numerous organisms have specialized organs to house their microbiota, yet the evolutionary processes underlying the origin of these tissues are not well understood (1–4). The Hawaiian bobtail squid, *Euprymna scolopes*, is a tractable animal model that has proven invaluable for interrogating symbiotic relationships and revealing common mechanisms by which successful colonization of animal epithelia by bacteria is established and maintained (Fig. 1*A*) (5). In the light organ (LO) of *E. scolopes*, a monospecific association with the bioluminescent bacterium *Vibrio fischeri* provides camouflage for the host and triggers a coordinated developmental remodeling of the surrounding epithelial tissues (6). Additionally, *E. scolopes* harbors a complex microbial consortium in the accessory nidamental gland (ANG), a component of the female reproductive system of many squids and cuttlefish that is hypothesized to play a role in egg defense (Fig. 1*B*) (7, 8). *Octopus bimaculoides*, the other cephalopod for which a genome has been reported to date, has neither a light organ nor an ANG making it an ideal organism for comparison (Fig. 1*C*). Therefore, to define elements of genome structure and function that are critical for symbiosis and the evolution of cephalopods, we sequenced and characterized the *E. scolopes* genome, the first example from the superorder Decapodiformes.

Results and Discussion

Our sequencing efforts have resulted in the most comprehensive cephalopod genome assembly to date, revealing the expansive and highly repetitive nature of the *E. scolopes* genome. Using a hybrid approach of shotgun and long-range linkage sequencing methods, an assembly estimated to be 5.1 Gigabases (Gb) was generated with a total half of the assembly in scaffolds of 3.7 Mb or longer (N50). The genome annotation was guided by 31 RNA-seq (Illumina) and ISO-seq (PacBio) transcriptomic libraries prepared from diverse tissues and developmental stages (*SI Appendix*, Table S1). The annotation resulted in 29,259 expressed

Significance

Animal–microbe associations are critical drivers of evolutionary innovation, yet the origin of specialized symbiotic organs remains largely unexplored. We analyzed the genome of *Euprymna scolopes*, a model cephalopod, and observed large-scale genomic reorganizations compared with the ancestral bilaterian genome. We report distinct evolutionary signatures within the two symbiotic organs of *E. scolopes*, the light organ (LO) and the accessory nidamental gland (ANG). The LO evolved through subfunctionalization of genes expressed in the eye, indicating a deep evolutionary link between these organs. Alternatively, the ANG was enriched in novel, species-specific orphan genes suggesting these two tissues originated via different evolutionary strategies. These analyses represent the first genomic insights into the evolution of multiple symbiotic organs within a single animal host.

Author contributions: M.B., M.M.-N., O.S., J.S.F., and S.V.N. designed research; M.B., G.C., S.J.M., A.M.S., S.M.-G., M.S.P., T.H.O., N.K., E.J.K., A.J.C., H.N., S.L., I.G.-F., P.M., E.S., G.W., D.S.R., O.S., J.S.F., and S.V.N. performed research; S.M.-G., M.S.P., T.H.O., N.K., E.J.K., A.J.C., H.N., S.L., P.M., E.S., G.W., D.S.R., M.M.-N., O.S., J.S.F., and S.V.N. contributed new reagents/analytic tools; M.B., G.C., S.J.M., H.S., A.M.S., H.N., S.L., P.M., E.S., G.W., D.S.R., O.S., J.S.F., and S.V.N. analyzed data; and M.B., G.C., S.J.M., H.S., A.M.S., S.M.-G., M.S.P., T.H.O., N.K., E.J.K., A.J.C., H.N., S.L., I.G.-F., P.M., E.S., G.W., D.S.R., M.M.-N., O.S., J.S.F., and S.V.N. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Genome and transcriptome sequencing reads have been deposited in the National Center for Biotechnology Information Sequence Read Archive, https://www.ncbi.nlm.nih.gov/sra (BioProject PRJNA470951).

[1]To whom correspondence may be addressed. Email: oleg.simakov@univie.ac.at, jfoster@ufl.edu, or spencer.nyholm@uconn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1817322116/-/DCSupplemental.

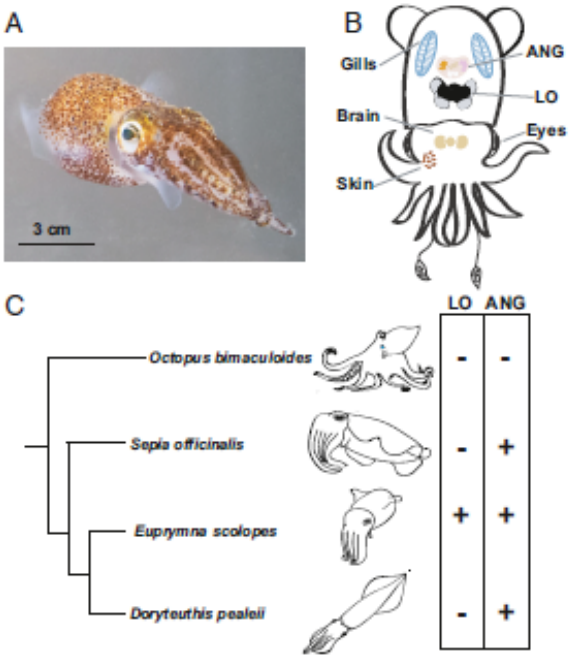


Fig. 1. The Hawaiian bobtail squid, *Euprymna scolopes*, a model host for microbiome research and cephalopod innovations. (*A*) The animal, shown here in the water column, is a nocturnal predator that uses the luminescence of the LO symbiont *Vibrio fischeri* for camouflage. Scale bar, 3 cm. Image courtesy of Elizabeth Ellenwood (photographer). (*B*) Overview of key symbiotic and nonsymbiotic organs within *E. scolopes*. Whereas males only have an LO symbiosis, the females have the additional symbiosis of the ANG, a reproductive organ housing a consortium of bacteria from predominantly two phyla (8). In addition to these symbiotic tissues, gene expression was also analyzed from the gills, brain, eye, and skin. (*C*) Distribution of LOs and ANGs that occur only in coleoid cephalopods. Branch lengths are derived from Tanner et al. (35).

protein-coding genes that yielded a BUSCO (Benchmarking Universal Single-Copy Orthologs) (9) score of 97% against the eukaryotic core set (*Materials and Methods*). A table summarizing the key statistics and a comparison with the *O. bimaculoides* genome is provided in *SI Appendix*, Table S2.

Genome analysis revealed differing histories of repeat element expansions between the bobtail squid and octopus, which may have contributed to the observed divergence in genome size (*SI Appendix*, Fig. S1). Although the estimated proportion of repetitive elements was >50% in both species, the abundance within individual repetitive element classes was strikingly different across the two species. Specifically, the abundance of long and short interspersed nuclear elements (LINEs and SINEs, respectively) differed; LINEs were the most abundant repeat class in *E. scolopes*, whereas SINEs dominated in *O. bimaculoides* (*SI Appendix*, Fig. S1*B*). Gene families in *E. scolopes* are similar to those reported in *O. bimaculoides* and also include tandem expansions in protocadherins and zinc finger transcription factors (*SI Appendix*, Fig. S2) (10). We infer that the last common coleoid cephalopod ancestor had 7,650 gene families, similar to the previously reported estimates for spiralian and bilaterian ancestors (10, 11). Because we do not find signatures of whole-genome duplication or evidence of extensive horizontal gene transfer in *E. scolopes*, the difference in the genome sizes between octopus and bobtail squid can be attributed to the increased LINE content in the *E. scolopes* genome. Future sequencing of other Decapodiformes genomes should enable a more exact estimate of this repeat element expansion.

The *E. scolopes* genome revealed key genomic transitions toward cephalopod genomic architecture. We find that over half of the local gene linkages (i.e., microsynteny) conserved across many noncephalopod bilaterian genomes were disrupted in both *O. bimaculoides* and *E. scolopes*, indicating that a large genomic reorganization took place in the cephalopod ancestor (Fig. 2*A*). Additionally, we found numerous linkages shared between octopus and the bobtail squid that have not been previously identified in other animal genomes (Fig. 2 *B* and *C*). Those linkages contain genes expressed in highly developed organ systems (e.g., the central nervous system) as well as in symbiotic organs from *E. scolopes* and testes of *O. bimacuolides* (Fig. 2*C*). In total, this taxon-specific microsynteny constituted 67% (126 out of 189) of all conserved syntenic linkages in both cephalopod genomes (*SI Appendix*, Fig. S3 and Table S2).

The improved contiguity of the genome assembly also uncovered the presence of a partial *hox* cluster in *E. scolopes* with very large (1.5–2 Mb) intergenic distances that is in stark contrast to the previously reported atomized *hox* cluster in the octopus genome (10). These are to our knowledge the longest inter *hox*-gene separations (Fig. 2*D*) known in animals and are consistent with the long, gene-free regions surrounding *hox* genes in *O. bimaculoides* (10) and the identified expansion of repetitive elements. Such large intergenic distances may be responsible for the evolution of unique regulatory mechanisms. This hypothesis is further supported by unconventional *hox* gene expression in *E. scolopes* (12).

To understand how this genome reorganization drove the evolution of symbiotic organs in *E. scolopes*, we studied genomic signatures of genes with tissue-specific expression in symbiotic versus nonsymbiotic organs. To identify possible candidate genes that evolved in symbiotic organs after a duplication event, we searched the *E. scolopes* genome for all paralogous pairs of genes in which one gene had tissue-specific expression and the second did not. Using these genes, we then calculated the number of synonymous substitutions (dS) between each gene pair (Fig. 3). We found that for the LO and ANG, the dS distance between tissue-specific genes and the respective closest non-LO and non-ANG paralog was relatively small [dS < 1, ~130 Mya, using a calibration point from Albertin et al. (10)]. These distances were in sharp contrast compared with genes specific to the eye and brain, which were much older (dS > 2, squid–octopus divergence, ~270 Mya) (10, 13). These paralog ages are consistent with a scenario in which the ANG and LO, which harbor dense populations of bacteria, underwent relatively recent innovations within the bobtail squid lineage through gene duplication, compared with the more ancient duplications of genes that are involved in development and function of cephalopod eyes and the nervous system.

The main mechanism behind gene duplication in *O. bimaculoides* is through tandem or segmental duplication (10). Genes exclusively expressed in either LO, eyes, or ANG were typically found in tandem clusters of paralogs located on single scaffolds. In particular, we found two gene clusters, the first composed of reflectins and the second of peroxidases that were expressed in the LO (see below, Fig. 4, and *SI Appendix*, Figs. S4–S6). In both cases, the tandem gene duplications forming those clusters were *E. scolopes*-specific and occurred after the octopus–squid split, consistent with the emergence of ANG and LO in the subsequent squid lineage. A large 2-Mb cluster of S-crystallins, which are predominantly expressed in the eyes (*SI Appendix*, Fig. S7), was expanded in tandem in *E. scolopes* but was missing from the octopus genome. The omega-crystallins, which are typically found in the LO and the eyes of *E. scolopes*, on the other hand, did not show signs of any expansion (*SI Appendix*, Fig. S7). The characteristic cluster in the ANG (~1 Mb)

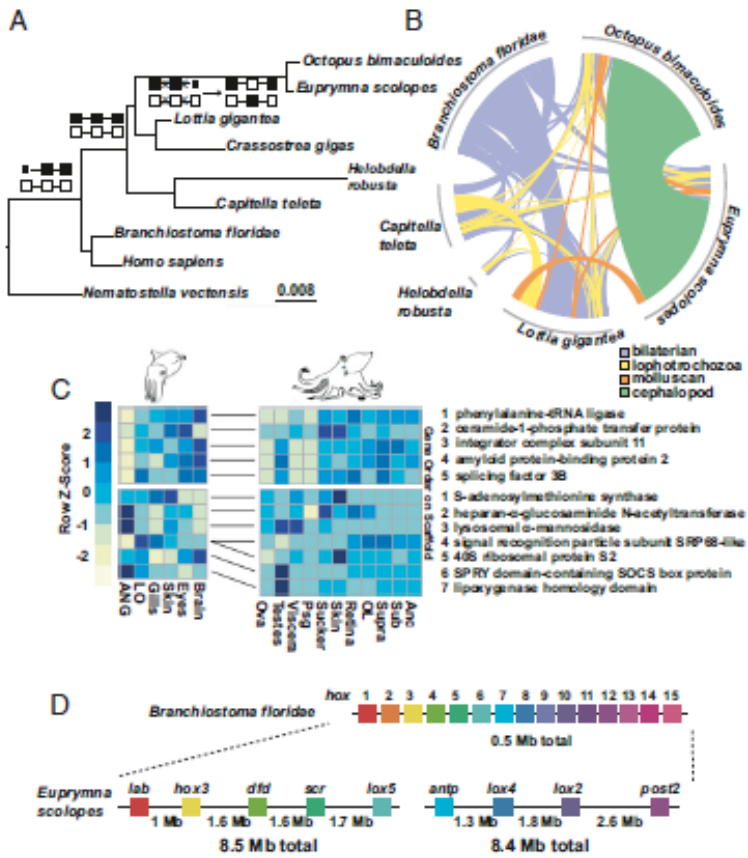


Fig. 2. Establishment of coleoid cephalopod genome architecture. (*A*) High rate of genome reorganization at the base of the coleoid cephalopods, as measured by the cumulative amount of microsynteries lost and gained. Branch length estimation using MrBayes (*SI Appendix*) on a fixed tree topology using binary presence or absence matrix of shared orthologous microsyntenic blocks. Squares (black/white) connected by lines above the nodes indicate two hypothetical microsyntenic blocks illustrating a common scenario where conserved bilaterian and lophotrochozoan synteny is disrupted (crossed lines) followed by the emergence of a new order through rearrangement in the cephalopod stem lineage. (*B*) Prevalence of the unique cephalopod microsynteny (green shaded area) in *O. bimaculoides* and *E. scolopes* genomes. Total length of arches for individual species corresponds to the number of genes in microsynteries. (*C*) Heat map of cephalopod unique microsyntenic clusters showing both neuronal (*Upper*) and broader gene expression (*Lower*) in *E. scolopes* and *O. bimaculoides*. Color bar indicates relative normalized gene expression level. Individual orthologous genes between *E. scolopes* and *O. bimaculoides* are connected by solid lines between heat maps with numbers indicating gene order along the scaffold. ANC, axial nerve cord; OL, optic lobe; Psg, posterior salivary gland; Supra/Sub, supraesophageal and subesophageal brain. (*D*) A large partial *hox* cluster on two separate scaffolds was recovered, totaling to a length of at least 16 Mb. *Branchiostoma floridae hox* cluster is shown for comparison with colors indicating orthologous genes.

contained both novel genes and transposable elements. Overall, there were significantly more tandemly duplicated genes expressed in the LO and eyes compared with the number of tandemly clustered genes specifically expressed in other tissues (Fisher's exact test $P < 0.001$), which underscores the importance of gene duplication for those organ systems.

Our data also reveal that despite the broad phylogenetic distribution among cephalopods (14), the ANG of *E. scolopes* is highly derived and shows an elevated evolutionary accumulation in both its coding (novel gene formation) and noncoding (turnover of the regulatory sequence) complements. The ANG is a secretory organ within the female reproductive system containing a bacterial consortium that is deposited into the egg capsule and is believed to play a role in defense from fouling and/or pathogens during embryogenesis (7, 8, 15, 16). Taxonomically restricted (i.e., orphan) genes have contributed to the evolution of unique tissues and organs in a number of animals (17–19). Within the two symbiotic organs of *E. scolopes* (LO and ANG), we profiled protein-coding transcripts that mapped to genome scaffolds from tissue-specific transcriptomes against the National Center for Biotechnology Information's nonredundant database [using BLASTP (Basic Local Alignment Search Tool - Protein) e-value threshold of 1E-5; Fig. 3*A*] and classified the transcripts as either *E. scolopes*-specific, cephalopod, molluscan, bilaterian, metazoan, or premetazoan (*SI Appendix*, Fig. S8).

The ANG had the highest proportion of *E. scolopes*-specific transcripts (>35%) among all tissues. Additionally, regions 20 kb up- and down-stream of genes specifically expressed in the ANG also showed a higher proportion of repetitive element content compared with genes expressed in any other tissue (Fig. 3*B*), suggesting a high evolutionary turnover in regulatory regions. These higher rates of genomic innovation, specific to the *E. scolopes* ANG genes, are also supported by our analysis of previously reported ANG transcriptomes from both *E. scolopes* and the (loliginid) swordtip squid, *Uroteuthis edulis*, the latter of

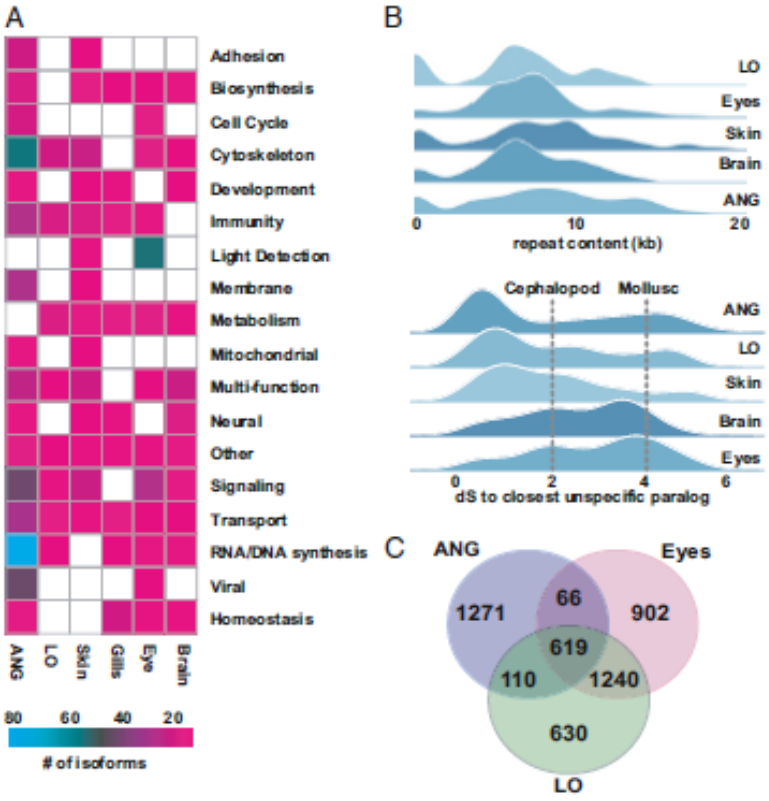


Fig. 3. Characterization (i.e., functional categories) of key tissues revealing the high contribution of novel genes toward ANG evolution as well as strong similarity between LO and eye transcriptomes. (*A*) Total counts of unique isoforms across different functional categories in six adult tissues. (*B*) (*Upper*) Joy plot of the number of nucleotides within 20 kilobases (kb) windows located up and downstream of the tissue-specific genes from (*A*) that are attributed to repetitive elements. Regions around ANG genes show higher repeat content compared with other tissues ($P < 0.1$, Wilcoxon rank sum test). (*Lower*) Joy plot of the synonymous substitutions distances (dS) between the genes specifically expressed in a given tissue to their closest paralog expressed elsewhere. Distributions from tissues representing cephalopod synapomorphies (brain and eyes) show an older mean ($P < 0.1$, Wilcoxon rank sum test) compared with the distributions from ANG, LO, and skin tissues. (*C*) Venn diagram representing the number of shared transcripts among LO, ANG, and eye tissues identifying a significant overlap between LO and eye transcripts.

which had fewer unique genes specific to this organ (*SI Appendix*, Fig. S8) (20).

Within the LO of *E. scolopes*, there were several other genetic signatures contributing to the distinctive features of this symbiotic organ. First, the dominant transcripts within the light organ were reflectins (Fig. 4*A*), providing strong support for previous studies indicating that the anatomical features used to modulate light in the LO are similar to the eye, as are their physiology, biochemistry, molecular biology, and developmental induction (21–23). A comparison of adult LO and eye transcriptomes supports these previous studies (Figs. 3*C* and 4*A*). Identified reflectins (Fig. 4*A*) were almost exclusively (>99.99% of total expression) expressed in the LO, eyes, and skin. LO expression constituted around 19.5% of the total reflectin expression, whereas eyes and skin composed 72.8% and 7.6%, respectively. Reflectins usually shared two or more expression domains and were rarely expressed in a single tissue. The majority of genes showed a distinctive LO and eyes expression domain, and no reflectins were expressed exclusively in both LO and skin (Fig. 4 *A* and *C*). Evolutionary timing of the LO–eye reflectin clusters may thus hint at the origin of the LO. Molecular dating analysis using branch length and squid–octopus split as a calibration point (10, 13) revealed the age of those clusters at around 30 Mya or younger. Additionally, among the genomic expansion of 14 peroxidases unique to *E. scolopes* (Fig. 4 *B* and *C* and *SI Appendix*, Fig. S6), six genes [including a previously reported halide peroxidase (24, 25)] formed a single genomic cluster expressed exclusively in the light organ. Unlike reflectins, this expansion did not form a monophyletic group, suggesting high gain or loss of genes within this family, or gene conversion. The host is known to produce a halide peroxidase that is expressed in the LO and generates hypohalous acid, a potent antimicrobial compound thought to contribute to specificity in the squid–vibrio association (26, 27). This peroxidase, along with several others, is expressed in the tissues directly in contact with *V. fischeri* (28, 29). The expansion of the peroxidases and the unique expression of a subset of these genes exclusively in the LO may reflect the development of a unique microenvironment that helps to maintain specificity with *V. fischeri* (*SI Appendix*, Fig. S6).

Analysis of protein families also revealed a number of genes that had multiple immune-associated domains, many of which were expressed in the LO, ANG, and skin (*SI Appendix*, Fig. S9). All three of these tissues interact directly with environmental and/or symbiotic microorganisms. A number of cellular and biochemical components of the innate immune system are known to play critical roles in mediating specificity in the LO symbiosis with *V. fischeri* (30–32). Previous observations of host hemocytes infiltrating the ANG (7), along with expression of immune-related genes reported here, suggest that the host's immune system plays a critical role in this symbiotic organ as well.

Although bacterial LOs are restricted to just two families of cephalopods (sepiolids and loliginids), the ANG is a more broadly distributed organ, present in these and a number of other squid and cuttlefish species (Fig. 1*C*). Their distinct evolutionary histories and origins are supported by the different evolutionary patterns detected. The prevalence of functional novelty through gene duplication in the LO compared with taxon-specific genes in the ANG suggests two different patterns of genomic innovation underlying symbiotic organ evolution in *E. scolopes*. Because these two associations are not nutritionally coupled, the mechanisms for the evolution of these symbiotic organs may be different from what has been proposed for insect bacteriocytes where metabolic complementation appears

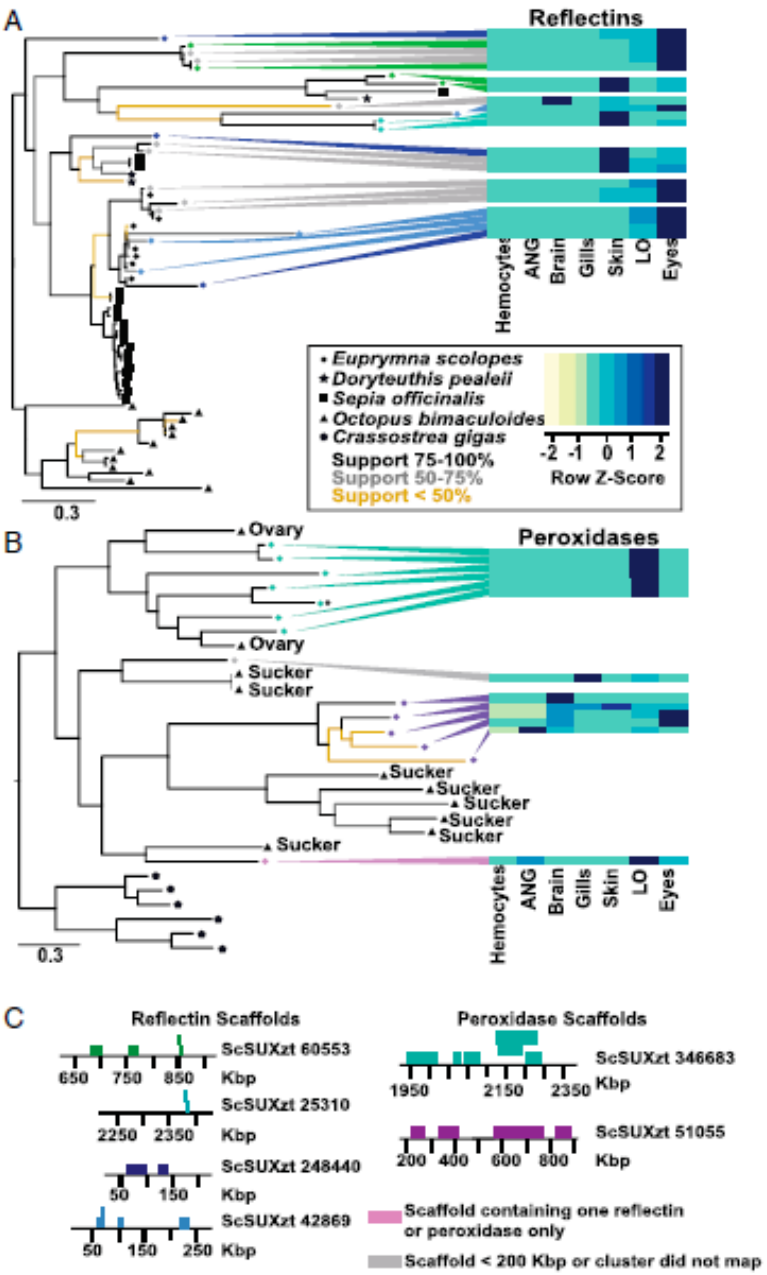


Fig. 4. Independent tandem gene cluster formation in squid and octopus and the origin of light organ-specific gene expression. Phylogenetic trees highlight highly specific expansion patterns that correlate with general shared expression between the light organ and the eyes for reflectins (*A*) or the appearance of the light organ-specific expression pattern for heme peroxidases (*B*). Heat maps indicate relative normalized expression levels for each tissue (*Z* scores). Scale bar underneath phylogenetic trees indicates amino acid substitutions per site. Color of the nodes (*A* and *B*) identifies genomically colocalized genes (shown in *C*). Those genes are clustered in tandem on several scaffolds with positions and gene identifiers labeled. (*C*) Tandem clusters of reflectin and peroxidase genes in the *E. scolopes* genome. Scaffold ID and approximate location (kbp) is shown for each gene (represented by rectangle). Colors correspond to the sequences on the trees in *A* for reflectins and (*B*) for peroxidases.

to be the major driver of innovation (33). Rapid gene evolution and/or gene loss in ancestral lineages may also have contributed to these unique signatures, especially in the ANG. Together, the presence of these two distinctive tissues within *E. scolopes* renders it a unique model organism to investigate the genetic mechanisms associated with the evolution of symbiotic organs (34).

Conclusions

Our analyses of the Hawaiian bobtail squid genome revealed large-scale genome reorganization, which preceded the coleoid cephalopod radiation, and provide evidence for two different patterns of genomic evolution that contributed to functional novelty. First, the extensive structural reorganization through the loss of ancient bilaterian microsynteny and increase in genome size through repetitive element expansions resulted in a unique genomic architecture that may have contributed to the innovations in the general cephalopod body plan. Second, the coding and surrounding regulatory regions of microbe-associated tissues suggest distinctive evolutionary patterns, such as the expansion of genes associated with immunity and light production in the LO. In the ANG, the prevalence of taxon-specific, or orphan, gene expression suggests a highly derived organ that evolved by a different means than the LO, perhaps due to unique selection

pressures. Overall, our results set the stage for further functional analysis of genomic innovations that led to the evolution of symbiotic organs in a morphologically and behaviorally complex clade.

Materials and Methods

Data Access. Genome and transcriptome sequencing reads have been deposited in the Sequence Read Archive as Bioproject PRJNA470951.

Genome Sequencing and Assembly. Genomic DNA generated in this study was derived from a single adult male *E. scolopes* (*SI Appendix*). Illumina reads from several libraries (*SI Appendix*, Table S1) were assembled with Meraculous resulting in a preliminary assembly; scaffolds were generated using the in vitro chromatin conformation capture (Chicago) at Dovetail Genomics.

Gene Model Predictions. Gene prediction models were generated using Augustus training with transcriptomic data. The prediction was filtered to exclude sequences that overlap with masked repeats over at least 50% of the lengths. From the remaining models, those with support from transcript evidence and significant homology to metazoan protein sequences were included in the nonredundant database. This gene modeling set was combined with transcriptome mapping (*SI Appendix*, Table S2) resulting in 29,089 protein sequences that passed the filtering stage.

Transcriptome Sequencing. Thirty-one transcriptomes of *E. scolopes* were included in the reference transcriptome and sample metadata of the tissues and time points is provided in *SI Appendix*, Table S1. Different RNA extractions and sequencing platforms were used and are listed in *SI Appendix*, Table S1. For the tissue-specific transcriptomes, RNA was extracted from ANG, brain, eyes, gills, hemocytes, LO, and skin tissues, as well as juvenile head (white body, optic nerve, and brain), eyes, gills, and light organ. For the PacBio IsoSeq library, RNA was extracted, normalized, and pooled from adult brain, eyes, white body, optical lobe, gills, LO, and skin as well as whole juvenile hatchling, 24-h aposymbiotic, and 24-h symbiotic animals (*SI Appendix*, Table S1).

Synteny Analysis. Using previously published methods of phylogeny-informed clustering, we have constructed the sets of orthologous gene families between the following species (11): *Capitella teleta*, *Helobdella robusta*, *Lottia gigantea*, *Octopus bimaculoides*, *Euprymna scolopes*, *Crassostrea gigas*, *Nematostella vectensis*, and *Branchiostoma floridae*. To account for differential gene loss that may impede our quantification of synteny loss/gain we focused only on 3,547 clusters that have an ortholog from each of the species. We have implemented microsynteny detection algorithm, as described in Simakov et al. (11), and found, in accordance with previous results (10, 34), ~600 microsyntenic blocks (in 163 orthologous groups) that can be traced back to the bilaterian ancestor because they are shared between either both ingroups (protostome and deuterostome) or an ingroup and an outgroup species (*Nematostella*) (Dataset S1).

ACKNOWLEDGMENTS. We thank Bo Reese from the Center for Genome Innovation, University of Connecticut (UCONN), for sequencing assistance and Jill Wegrzyn, Stephen King, and Ion Moraru of the Computational Biology Core, Institute for Systems Genomics, UCONN, for computational support. Computation was conducted at the Vienna Life Sciences Cluster (cube.univie.ac.at/) and the computing cluster at the Okinawa Institute of Science and Technology (OIST). This work was supported in part by NIH Grant R01-AI50661 (to M.M.-N. and E. G. Ruby, co-Principal Investigator); NIH Grant R01-OD11024 (to E. G. Ruby, co-Principal Investigator, and M.M.-N.); the University of Wisconsin (M.M.-N.); NASA Space Biology Grant NNX13AM44G (to J.S.F.); NSF Integrative Organismal Systems 1557914 and the Office of the Vice President for Research, UCONN (to S.V.N.); and the Molecular Genetics Unit of OIST Graduate University (to D.S.R.). O.S. and H.S. were supported by a grant from the Austrian Science Fund (Fonds zur Förderung der Wissenschaftlichen Forschung) P30686-B29.

1. Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42:165–190.
2. Markmann K, Parniske M (2009) Evolution of root endosymbiosis with bacteria: How novel are nodules? *Trends Plant Sci* 14:77–86.
3. Matsuura Y, et al. (2012) Evolution of symbiotic organs and endosymbionts in lygaeid stinkbugs. *ISME J* 6:397–409.
4. Griesmann M, et al. (2018) Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* 361:eaat1743.
5. McFall-Ngai MJ (2014) The importance of microbes in animal development: Lessons from the squid-vibrio symbiosis. *Annu Rev Microbiol* 68:177–194.
6. McFall-Ngai MJ (2015) Giving microbes their due—Animal life in a microbially dominant world. *J Exp Biol* 218:1968–1973.
7. Collins AJ, et al. (2012) Diversity and partitioning of bacterial populations within the accessory nidamental gland of the squid *Euprymna scolopes*. *Appl Environ Microbiol* 78:4200–4208.
8. Kerwin AH, Nyholm SV (2017) Symbiotic bacteria associated with a bobtail squid reproductive system are detectable in the environment, and stable in the host and developing eggs. *Environ Microbiol* 19:1463–1475.
9. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
10. Albertin CB, et al. (2015) The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524:220–224.
11. Simakov O, et al. (2013) Insights into bilaterian evolution from three spiralian genomes. *Nature* 493:526–531.
12. Lee PN, Callaerts P, De Couet HG, Martindale MQ (2003) Cephalopod Hox genes and the origin of morphological novelties. *Nature* 424:1061–1065.
13. Kröger B, Vinther J, Fuchs D (2011) Cephalopod origin and evolution: A congruent picture emerging from fossils, development and molecules: Extant cephalopods are younger than previously realised and were under major selection to become agile, shell-less predators. *BioEssays* 33:602–613.
14. Lindgren AR, Pankey MS, Hochberg FG, Oakley TH (2012) A multi-gene phylogeny of Cephalopoda supports convergent morphological evolution in association with multiple habitat shifts in the marine environment. *BMC Evol Biol* 12:129.
15. Biggs J, Epel D (1991) Egg capsule sheath of *Loligo opalescens* Berry: Structure and association with bacteria. *J Exp Zool* 259:263–267.
16. Barbieri E, et al. (2001) Phylogenetic characterization of epibiotic bacteria in the accessory nidamental gland and egg capsules of the squid *Loligo pealei* (Cephalopoda: Loliginidae). *Environ Microbiol* 3:151–167.
17. Gould RM, et al. (2008) Myelin sheaths are formed with proteins that originated in vertebrate lineages. *Neuron Glia Biol* 4:137–152.
18. Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TC (2009) More than just orphans: Are taxonomically-restricted genes important in evolution? *Trends Genet* 25:404–413.
19. Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. *Nat Rev Genet* 12:692–702.
20. Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH (2014) Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. *Proc Natl Acad Sci USA* 111:E4736–E4742.
21. Tong D, et al. (2009) Evidence for light perception in a bioluminescent organ. *Proc Natl Acad Sci USA* 106:9836–9841.
22. McFall-Ngai M, Heath-Heckman EA, Gillette AA, Peyer SM, Harvie EA (2012) The secret languages of coevolved symbioses: Insights from the *Euprymna scolopes-Vibrio fischeri* symbiosis. *Semin Immunol* 24:3–8.
23. Crookes WJ, et al. (2004) Reflectins: The unusual proteins of squid reflective tissues. *Science* 303:235–238.
24. Tomarev SI, et al. (1993) Abundant mRNAs in the squid light organ encode proteins with a high similarity to mammalian peroxidases. *Gene* 132:219–226.
25. Weis VM, Small AL, McFall-Ngai MJ (1996) A peroxidase related to the mammalian antimicrobial protein myeloperoxidase in the *Euprymna-Vibrio* mutualism. *Proc Natl Acad Sci USA* 93:13683–13688.
26. Visick KL, Ruby EG (1998) The periplasmic, group III catalase of *Vibrio fischeri* is required for normal symbiotic competence and is induced both by oxidative stress and by approach to stationary phase. *J Bacteriol* 180:2087–2092.
27. Small AL, McFall-Ngai MJ (1999) Halide peroxidase in tissues that interact with bacteria in the host squid *Euprymna scolopes*. *J Cell Biochem* 72:445–457.
28. Wier AM, et al. (2010) Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. *Proc Natl Acad Sci USA* 107:2259–2264.
29. Schleicher TR, Nyholm SV (2011) Characterizing the host and symbiont proteomes in the association between the bobtail squid, *Euprymna scolopes*, and the bacterium, *Vibrio fischeri*. *PLoS One* 6:e25649.
30. McFall-Ngai M, Nyholm SV, Castillo MG (2010) The role of the immune system in the initiation and persistence of the *Euprymna scolopes-Vibrio fischeri* symbiosis. *Semin Immunol* 22:48–53.
31. McAnulty SJ, Nyholm SV (2017) The role of hemocytes in the Hawaiian bobtail squid, *Euprymna scolopes*: A model organism for studying beneficial host-microbe interactions. *Front Microbiol* 7:2013.
32. Chen F, et al. (2017) Bactericidal permeability-increasing proteins shape host-microbe interactions. *MBio* 8:e00040-17.
33. Wilson AC, Duncan RP (2015) Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proc Natl Acad Sci USA* 112:10255–10261.
34. Simakov O, Kawashima T (2017) Independent evolution of genomic characters during major metazoan transitions. *Dev Biol* 427:179–192.
35. Tanner AR, et al. (2017) Molecular clocks indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic marine revolution. *Proc Biol Sci* 284:20162818.

Appendix III: ANG metagenome co-assembly antiSMASH results

Cluster type	Cluster number	Contig	Does the cluster cover the entire length of the contig/is cut off?	Contig taxonomy	Top homologous cluster hit	Percent of genes similar to top homologous hit	Most similar known cluster	Percent of genes similar to known cluster
Acyl amino acids	180	k141_120365	no	c_Alphaproteobacteria	Chelativorans sp. BNC1	9		
Arylpolyene	70	k141_39235	no	s__Nitrobacter_hamburgensis	Mesorhizobium sp. L48C026A00	8		
Arylpolyene	84	k141_46281	beginning of contig	NA	Flagellimonas sp. DK169	7		
Arylpolyene	98	k141_56034	beginning of contig	NA	Stenotrophomonas maltophilia EPM1	11	Aryl polyene carboxylic acid (Vibrio fischeri)	15
Arylpolyene	104	k141_61276	no	s__Vibrio_tubiashii	Photobacterium leiognathi KNH6	46	Aryl polyene carboxylic acid (Vibrio fischeri)	85
Arylpolyene	127	k141_80396	end of contig	s__Mesorhizobium_loti	Mesorhizobium sp. L48C026A00	7		
Arylpolyene	179	k141_119388	nearly entire contig	NA	Chryseobacterium piperi CTM	16	Flexirubin	11
Arylpolyene	206	k141_144372	no	NA	Flagellimonas sp. DK169	9		

Arylpolyene	211	k141_149194	no	NA	Verrucomicrobia bacterium IMCC26134	7		
Arylpolyene-resorcinol	87	k141_47931	no	NA	Chryseobacterium piperi CTM	20	Flexirubin	19
Bacteriocin	2	k141_1170	no	NA	No significant ClusterBlast hits found	NA		
Bacteriocin	16	k141_5178	no	NA	Shewanella woodyi ATCC 51908	6		
Bacteriocin	19	k141_7064	no	NA	Burkholderia oklahomensis EO147	4		
Bacteriocin	28	k141_12475	cut off at beginning	s__Phaeobacter_gallaeciensis	Ruegeria sp. ANG-R	15		
Bacteriocin	35	k141_17677	nearly entire contig	s__Phaeobacter_inhibens	Leisingera sp. ANG-M6	6		
Bacteriocin	36	k141_18014	no	s__Leisingera_methylohalidivorans	Leisingera sp. ANG1	9		
Bacteriocin	38	k141_19264	no	s__Ruegeria_pomeroyi	Ruegeria sp. ANG-S4	38		
Bacteriocin	47	k141_27164	nearly entire contig	p__Proteobacteria	Shimia marina CECT7688	8		
Bacteriocin	49	k141_27712	no	NA	No significant ClusterBlast hits found	NA		
Bacteriocin	58	k141_31748	no	s__Vibrio_tubiashii	Vibrio tubiashii ATCC 19109	31		
Bacteriocin	59	k141_32624	no	s__Pelagibaca_abyssi	Ruegeria sp. ANG-R	9		

Bacteriocin	63	k141_33674	no	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-M1	20		
Bacteriocin	64	k141_34511	no	s__Ruegeria_mobilis	Ruegeria sp. ANG-S4	17		
Bacteriocin	66	k141_35480	cut off at beginning	s__Phaeobacter_gallaeciensis	Sulfitobacter pontiacus 3SOLIMAR09	20		
Bacteriocin	75	k141_42149	end of contig	s__Pseudomonas_fluorescens	Shewanella sediminis HAW-EB3	4		
Bacteriocin	76	k141_42783	beginning of contig	s__Phaeobacter_inhibens	Leisingera sp. ANG-Vp	8		
Bacteriocin	78	k141_43109	end of contig	s__Erythrobacter_flavus	Sphingobium japonicum BiD32	23		
Bacteriocin	82	k141_44350	no	s__Ruegeria_pomeroyi	Ruegeria pomeroyi DSS-3	13		
Bacteriocin	88	k141_48547	no	s__Phaeobacter_inhibens	Leisingera sp. ANG-DT	6		
Bacteriocin	94	k141_53536	no	p__Proteobacteria	Ruegeria sp. CECT 5091	8		
Bacteriocin	95	k141_54860	beginning of contig	NA	Asticcacaulis sp. AC466	7		
Bacteriocin	114	k141_70470	beginning of contig	s__Phaeobacter_inhibens	Leisingera methylohalidivorans DSM 14336 MB2	6		
Bacteriocin	116	k141_71878	no	s__Tateyamaria_omphalii	Tateyamaria sp. ANG-S1	19		
Bacteriocin	117	k141_72976	beginning of contig	s__Phaeobacter_inhibens	Leisingera sp. ANG-DT	4		
Bacteriocin	136	k141_88169	beginning of contig	s__Phaeobacter_inhibens	Leisingera sp. ANG-Vp	4		

Bacteriocin	140	k141_92758	nearly entire contig	NA	Teredinibacter turnerae T7901	5		
Bacteriocin	148	k141_95477	no	NA	Flavobacterium sp. Root901	16		
Bacteriocin	149	k141_97332	no	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-DT	22		
Bacteriocin	150	k141_98387	beginning of contig	s__Antarctobacter_heliothermus	Sulfitobacter pseudonitzschiae H3	8		
Bacteriocin	155	k141_101463	end of contig	s__Phaeobacter_inhibens	Leisingera sp. ANG-Vp	6		
Bacteriocin	158	k141_104149	no	s__Leisingera_methylohalidivorans	Ruegeria atlantica CECT4292	31		
Bacteriocin	162	k141_107396	nearly entire contig	NA	Vibrio maritimus	5		
Bacteriocin	175	k141_116421	no	s__Vibrio_shilonii	Vibrio maritimus	20		
Bacteriocin	176	k141_116764	no	s__Alteromonas_macleodii	Cellvibrio sp. BR	10		
Bacteriocin	183	k141_121093	no	s__Ruegeria_sp._TM1040	Ruegeria atlantica CECT4292	27		
Bacteriocin	188	k141_122964	no	s__Tateyamaria_omphalii	Roseovarius indicus B108	14		
Bacteriocin	216	k141_155734	nearly entire contig	f__Rhodobacteraceae	Ruegeria pomeroyi DSS-3	4		
Bacteriocin	235	k141_170096	no	NA	Pseudogulbenkiania ferrooxidans 2002	8		
Bacteriocin	236	k141_172980	end of contig	s__Phaeobacter_gallaeciensis	Sulfitobacter pontiacus 3SOLIMAR09	17		

Bacteriocin	249	k141_179764	nearly entire contig	s__Hoeflea_sp._IMCC20628	No significant ClusterBlast hits found	NA		
Bacteriocin	252	k141_184734	no	s__Ruegeria_pomeroyi	Ruegeria atlantica CECT4292	20		
Bacteriocin	253	k141_184734	cut off at beginning	s__Ruegeria_pomeroyi	Leisingera sp. ANG-Vp	26		
Bacteriocin	254	k141_184734	no	s__Ruegeria_pomeroyi	Shimia sp. SK013	6		
Bacteriocin-lantipeptide	159	k141_104896	no	s__Phaeobacter_gallaeciensis	Thioclava sp. DT23-4	7		
Cyanobactin	39	k141_20287	end of contig	p__Bacteroidetes	No significant ClusterBlast hits found	NA		
Ectoine	29	k141_14222	no	c__Alphaproteobacteria	Ruegeria sp. ANG-R	10		
Ectoine	46	k141_27146	no	s__Celeribacter_manganoxidans	Maritimibacter alkaliphilus HTCC2654	7	Ectoine	40
Ectoine	121	k141_74138	no	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-M1	27		
Ectoine	123	k141_75836	no	s__Leisingera_methylohalidivorans	Leisingera methylohalidivorans DSM 14336 MB2	18	Ectoine	100
Ectoine	146	k141_95172	no	f__Rhodobacteraceae	Ruegeria sp. ANG-R	10		
Ectoine	164	k141_108103	no	s__Streptomyces_sp._RTd22	Ruegeria atlantica CECT4292	10		
Ectoine	184	k141_121345	end of contig	f__Rhodobacteraceae	Ruegeria sp. CECT 5091	11		

Ectoine	199	k141_135169	no	s__Leisingera_methylohalidivorans	Leisingera methylohalidivorans DSM 14336 MB2	18	Ectoine	100
Ectoine	204	k141_140760	end of contig	s__Ruegeria_pomeroyi	Leisingera sp. ANG-Vp	13		
Ectoine	221	k141_159192	beginning of contig	NA	Bradyrhizobium sp. ORS278	5		
Ectoine	228	k141_162365	no	NA	Ruegeria sp. ANG-R	7		
Homoserine lactone	6	k141_2038	no	s__Phaeobacter_gallaeciensis	Ruegeria sp. CECT 5091	10		
Homoserine lactone	7	k141_2059	no	NA	Phaeobacter gallaeciensis DSM 26640 plasmid pGal G40	6		
Homoserine lactone	8	k141_2157	no	s__Ruegeria_pomeroyi	Silicibacter lacuscaerulensis ITI-1157	44		
Homoserine lactone	11	k141_3545	no	s__Tateyamaria_omphalii	Ruegeria sp. ANG-R	12		
Homoserine lactone	12	k141_3823	no	s__Roseobacter_denitrificans	Sulfitobacter noctilucae NB-68 NB68	41		
Homoserine lactone	13	k141_4123	no	s__Rhizobium_phaseoli	Mesorhizobium huakuii 7653R	15		
Homoserine lactone	17	k141_5505	no	NA	No significant ClusterBlast hits found	NA		
Homoserine lactone	20	k141_8377	no	NA	No significant ClusterBlast hits found	NA		

Homoserine lactone	24	k141_94 62	end of contig	s__Phaeobacter _piscinae	Ruegeria sp. ANG-S4	4		
Homoserine lactone	26	k141_11 415	no	s__Leisingera_ methylohalidivo rans	Leisingera sp. ANG-S5	21		
Homoserine lactone	27	k141_11 487	no	f__Rhodobacter aceae	Ruegeria sp. ANG-R	4		
Homoserine lactone	30	k141_14 805	no	f__Rhodobacter aceae	Ruegeria sp. ANG-S4	9		
Homoserine lactone	32	k141_15 797	no	s__Ruegeria_po meroyi	Ruegeria pomeroyi DSS-3	19		
Homoserine lactone	34	k141_16 920	no	s__Ruegeria_po meroyi	Ruegeria atlantica CECT4293	21		
Homoserine lactone	43	k141_24 881	end of contig	s__Antarctobact er_heliothermus	Ruegeria sp. ANG-S4	4		
Homoserine lactone	45	k141_25 899	no	s__Phaeobacter _gallaeciensis	Roseobacter sp. AzwK-3b	23		
Homoserine lactone	53	k141_29 610	no	s__Leisingera_ methylohalidivo rans	Leisingera methylohalidivora ns DSM 14336 MB2	22		
Homoserine lactone	54	k141_29 709	no	s__Ruegeria_po meroyi	Silicibacter lacuscaerulensis ITI-1157	30		
Homoserine lactone	57	k141_31 566	no	s__Pelagibaca_ abyssi	Ruegeria sp. CECT 5091	9		
Homoserine lactone	60	k141_32 672	no	s__Thioclava_n itratireducens	Phaeobacter sp. CECT 5382	22		
Homoserine lactone	62	k141_33 447	no	s__Ruegeria_po meroyi	Ruegeria sp. ANG-S4	41		
Homoserine lactone	65	k141_35 082	end of contig	NA	No significant ClusterBlast hits found	NA		

Homoserine lactone	68	k141_36768	no	s__Mesorhizobium_loti	Mesorhizobium opportunistum WSM2075	28		
Homoserine lactone	81	k141_44204	no	NA	No significant ClusterBlast hits found	NA		
Homoserine lactone	91	k141_51640	beginning of contig	NA	Hyphomicrobium denitrificans 1NES1	2		
Homoserine lactone	92	k141_52158	no	s__Mesorhizobium_amorphae	Mesorhizobium huakuii 7653R	13		
Homoserine lactone	107	k141_64454	beginning of contig	NA	Roseobacter sp. AzwK-3b	5		
Homoserine lactone	122	k141_74615	no	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-S5	10		
Homoserine lactone	129	k141_81519	nearly entire contig	s__Leisingera_methylohalidivorans	Ruegeria mobilis F1926	22		
Homoserine lactone	131	k141_84104	no	s__Maribacter_sp._MAR_2009_60	Nitrosomonas communis Nm2	2		
Homoserine lactone	134	k141_87827	end of contig	s__Marinovum_algicola	Leisingera sp. ANG-S5	6		
Homoserine lactone	135	k141_87914	no	s__Nitratireductor_basaltis	Roseobacter sp. AzwK-3b	8		
Homoserine lactone	138	k141_90983	beginning of contig	NA	No significant ClusterBlast hits found	NA		
Homoserine lactone	139	k141_91064	no	f__Rhodobacteraceae	Ruegeria sp. CECT 5091	17		

Homoserine lactone	141	k141_93079	no	s__Phaeobacter_gallaeciensis	Sulfitobacter geojensis MM-124	19		
Homoserine lactone	144	k141_94674	no	s__Phaeobacter_gallaeciensis	Ruegeria sp. ANG-R	27		
Homoserine lactone	145	k141_94957	no	s__Phaeobacter_piscinae	Rhodobacterales bacterium Y4I	11		
Homoserine lactone	151	k141_98710	no	d__Bacteria	Tateyamaria sp. ANG-S1	4		
Homoserine lactone	168	k141_109947	no	s__Kosakonia_oryzae	Ruegeria sp. ANG-R	8		
Homoserine lactone	171	k141_112745	no	s__Ruegeria_pomeroyi	Ruegeria sp. ANG-R	39		
Homoserine lactone	177	k141_117763	no	s__Phaeobacter_gallaeciensis	Phaeobacter sp. S26	7		
Homoserine lactone	178	k141_118401	no	s__Marinovum_algicola	Ruegeria sp. ANG-S4	34		
Homoserine lactone	181	k141_120732	no	NA	Thioalkalivibrio thiocyanoxidans ARh 4	6		
Homoserine lactone	187	k141_122381	end of contig	NA	Roseobacter litoralis Och 14	2		
Homoserine lactone	190	k141_123203	no	s__Phaeobacter_gallaeciensis	Ruegeria sp. ANG-S4	30		
Homoserine lactone	192	k141_124292	no	s__Phaeobacter_gallaeciensis	Phaeobacter gallaeciensis DSM 26640 plasmid pGal	6		
Homoserine lactone	198	k141_135153	no	s__Sulfitobacter_pseudonitzschiae	Ruegeria sp. ANG-S4	25		

Homoserine lactone	202	k141_137565	no	NA	No significant ClusterBlast hits found	NA		
Homoserine lactone	203	k141_137621	no	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-M1	28		
Homoserine lactone	205	k141_141073	end of contig	s__Jannaschia_sp._CCS1	Tateyamaria sp. ANG-S1	4		
Homoserine lactone	208	k141_146064	no	s__Ruegeria_pomeroyi	Ruegeria atlantica CECT4292	44		
Homoserine lactone	214	k141_151325	no	s__Leisingera_methylohalidivorans	Leisingera sp. ANG1	4		
Homoserine lactone	218	k141_158190	no	s__Roseobacter_denitrificans	Leisingera sp. ANG-DT	14		
Homoserine lactone	219	k141_158623	no	s__Phaeobacter_gallaeciensis	Leisingera sp. ANG-S5	10		
Homoserine lactone	222	k141_159815	no	g__Phaeobacter	Leisingera sp. ANG-S5	10		
Homoserine lactone	225	k141_162072	end of contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-S5	6		
Homoserine lactone	227	k141_162352	end of contig	s__Pelagibacterium_halotolerans	No significant ClusterBlast hits found	NA		
Homoserine lactone	231	k141_165807	beginning of contig	s__Burkholderia_cenocepacia	Mesorhizobium sp. L2C054A000	5		
Homoserine lactone	233	k141_168395	no	s__Pseudomonas_litoralis	Roseobacter sp. AzwK-3b	17		
Homoserine lactone	234	k141_169236	no	NA	Ruegeria sp. CECT 5091	6		

Homoserine lactone	237	k141_173081	no	s__Antarctobacter_heliothermus	No significant ClusterBlast hits found	NA		
Homoserine lactone	244	k141_177919	no	s__Sinorhizobium_fredii	Jannaschia aquimarina GSW-M26	7		
Homoserine lactone	248	k141_179043	no	s__Candidatus_Puniceispirillum_marinum	No significant ClusterBlast hits found	NA		
Homoserine lactone	255	k141_184898	beginning of contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-Vp	7		
NRPS	15	k141_5094	beginning of contig	s__Tateyamaria_omphalii	Jannaschia donghaensis CECT7802	21		
NRPS	41	k141_21426	no	s__Paenibacillus_odorifer	Archangium gephyra DSM 2261	7		
NRPS	48	k141_27489	entire contig	NA	No significant ClusterBlast hits found	NA		
NRPS	50	k141_27738	entire contig	s__Streptomyces_violaceusniger	No significant ClusterBlast hits found	NA		
NRPS	77	k141_42863	no	s__Acidipropionibacterium_acidipropionici	Opitutus terrae PB90-1	12		
NRPS	90	k141_50996	no	NA	Pseudomonas sp. CCOS 191	6		
NRPS	106	k141_62207	entire contig	s__Vibrio_coralliilyticus	No significant ClusterBlast hits found	NA		

NRPS	112	k141_69895	no	s__Leisingera_methylohalidivorans	Ruegeria sp. ANG-S4	43	Asukamycin	3
NRPS	113	k141_70397	no	NA	Paenibacillus alvei TS-15	15		
NRPS	118	k141_73304	entire contig	NA	No significant ClusterBlast hits found	NA		
NRPS	125	k141_77367	entire contig	NA	No significant ClusterBlast hits found	NA		
NRPS	128	k141_81248	entire contig	NA	No significant ClusterBlast hits found	NA		
NRPS	130	k141_81952	no	o__Rhizobiales	Beijerinckia indica subsp. indica ATCC 9039	10		
NRPS	132	k141_85662	entire contig	s__Vibrio_coralliilyticus	Nocardia farcinica NCTC11134	3		
NRPS	156	k141_103364	no	c__Gammaproteobacteria	Bacillus amyloliquefaciens G341	7		
NRPS	160	k141_105205	end of contig	s__Vibrio_tubiashii	Vibrio tubiashii ATCC 19109	22	Cupria-chelin	11
NRPS	172	k141_113535	entire contig	NA	No significant ClusterBlast hits found	NA		
NRPS	185	k141_121378	nearly entire contig	NA	Winogradskyella sp. PG-2	7		

NRPS	194	k141_126573	no	s__Thauera_sp._MZ1T	Nodularia spumigena CCY9414	6		
NRPS	217	k141_157404	no	s__Mesorhizobium_loti	Tistrella mobilis KA081020-065	23		
NRPS	220	k141_158838	no	c__Alphaproteobacteria	Paenibacillus alvei DSM 29	11		
NRPS	223	k141_160375	entire contig	NA	No significant ClusterBlast hits found	NA		
NRPS	224	k141_161014	nearly entire contig	s__Burkholderia_cepacia	Pseudomonas entomophila str. L48	2		
NRPS	229	k141_163112	end of contig	s__Lacunisphaera_limnophila	Cellvibrio japonicus Ueda107	10		
NRPS	240	k141_175343	nearly entire contig	NA	Bacillus methylotrophicus JJ-D34	9		
NRPS	245	k141_178601	no	s__Arenibacter_algicola	Winogradskyella sp. PG-2	7		
Nucleoside	74	k141_41991	no	NA	Streptosporangium amethystogenes	3		
Other	5	k141_2028	end of contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-Vp	21		
Other	51	k141_28750	no	s__Thiocystis_violascens	Syntrophobotulus glycolicus DSM 8271	8		
Other	52	k141_29169	entire contig	NA	No significant ClusterBlast hits found	NA		

Other	61	k141_32853	no	NA	No significant ClusterBlast hits found	NA		
Other	69	k141_38158	no	NA	No significant ClusterBlast hits found	NA		
Other	100	k141_57038	end of contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-S	21		
Other	101	k141_57453	beginning of contig	s__Ruegeria_mobilis	Celeribacter indicus P73	14		
Other	102	k141_58930	no	NA	Bartonella schoenbuchensis m07a	6		
Other	103	k141_59605	no	NA	No significant ClusterBlast hits found	NA		
Other	108	k141_65246	no	s__Leisingera_methylohalidivorans	Roseobacter sp. MED193	21	Scabichelin	20
Other	110	k141_66180	beginning of contig	s__Yangia_sp._CCB-MM3	Ruegeria sp. ANG-S4	10		
Other	119	k141_73955	no	NA	No significant ClusterBlast hits found	NA		
Other	124	k141_76498	end of contig	NA	No significant ClusterBlast hits found	NA		
Other	126	k141_79191	beginning of contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-M1	40		

Other	142	k141_93567	entire contig	NA	No significant ClusterBlast hits found	NA		
Other	147	k141_95334	beginning of contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-S	15		
Other	167	k141_109402	end of contig	s__Pelagibaca_abyssi	Loktanella sp. 5RATIMAR09	31		
Other	186	k141_121977	nearly entire contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-S	10		
Other	189	k141_123169	nearly entire contig	NA	No significant ClusterBlast hits found	NA		
Other	193	k141_125994	no	g__Streptomyces	Streptomyces sp. W007	7		
Other	196	k141_129950	no	s__Phaeobacter_gallaeciensis	Thioclava sp. DT23-4	9		
Other	200	k141_135291	end of contig	NA	Thalassobius mediterraneus	4		
Other	207	k141_144618	beginning of contig	NA	Microcoleus chthonoplastes PCC 7420	11		
Other	226	k141_162155	no	s__Shinella_sp._HZN7	Thalassobius sp. CECT 5113	5		
Other	242	k141_176308	end of contig	s__Streptomyces_sp._CFMR_7	Xanthomonas axonopodis pv. punicae str. LMG 859	11		
Other	250	k141_180214	nearly entire contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-M6	45		

Siderophore	79	k141_44198	end of contig	s__Novosphingobium_sp._PP1Y	Leisingera sp. ANG-DT	33		
Siderophore	154	k141_100117	entire contig	s__Leisingera_methylohalidivorans	No significant ClusterBlast hits found	NA		
Siderophore	166	k141_109362	entire contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-Vp	7		
Siderophore	173	k141_114330	nearly entire contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-M7	5		
Siderophore	191	k141_123609	nearly entire contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-S3	11		
Siderophore	195	k141_129343	nearly entire contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-S5	14		
Siderophore	210	k141_147557	nearly entire contig	d__Bacteria	Leisingera sp. ANG-DT	16		
T1PKS	23	k141_9118	entire contig	s__Phaeobacter_gallaeciensis	No significant ClusterBlast hits found	NA		
T1PKS	31	k141_15106	entire contig	s__Marinovum_algicola	No significant ClusterBlast hits found	NA		
T1PKS	40	k141_20322	entire contig	s__Leisingera_methylohalidivorans	No significant ClusterBlast hits found	NA		
T1PKS	80	k141_44202	beginning of contig	f__Rhodobacteraceae	Leisingera sp. ANG-Vp	13		

T1PKS	96	k141_55305	entire contig	NA	No significant ClusterBlast hits found	NA		
T1PKS	99	k141_56633	end of contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG1	39		
T1PKS	120	k141_74109	no	s__Labrenzia_sp._VG12	Neorhizobium galegae bv. Officinalis	16		
T1PKS	137	k141_88496	entire contig	s__Rhizobacter_gummiphilus	No significant ClusterBlast hits found	NA		
T1PKS	157	k141_104114	end of contig	s__Ruegeria_sp._TM1040	Loktanella sp. 1ANDIMAR09 PM06 003	20		
T1PKS	201	k141_136507	entire contig	s__Brevibacillus_laterosporus	No significant ClusterBlast hits found	NA		
T1PKS	209	k141_146608	nearly entire contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-DT	5		
T1PKS	215	k141_151472	nearly entire contig	s__Phaeobacter_gallaeciensis	Leisingera sp. ANG-Vp	8		
T1PKS	239	k141_174907	entire contig	s__Rhizobacter_gummiphilus	No significant ClusterBlast hits found	NA		
T1PKS	246	k141_178705	end of contig	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-M1	54		
T1PKS	247	k141_179019	entire contig	s__Pelagibaca_abyssi	No significant ClusterBlast hits found	NA		

T1PKS-NRPS	153	k141_99921	beginning of contig	g__Streptomyces	Phenylobacterium sp. Root77	4		
T1PKS-NRPS	197	k141_133102	entire contig	NA	No significant ClusterBlast hits found	NA		
T3PKS	14	k141_5077	end of contig	NA	Flagellimonas sp. DK169	7		
T3PKS	55	k141_30068	no	s__Tenacibaculum_sp._LPB0136	Flagellimonas sp. DK169	51		
T3PKS	97	k141_55541	no	NA	No significant ClusterBlast hits found	NA		
T3PKS	115	k141_70534	beginning of contig	NA	Oceaniovalibus guishaninsula JLT2003	4		
T3PKS-arylpolyene	21	k141_8533	no	s__Polaribacter_sp._KT25b	Zobellia galactanivorans DsiJT	44		
Terpene	1	k141_207	yes	NA	Sulfitobacter geojensis MM-124	5		
Terpene	3	k141_1241	no	s__Muricauda_ruestringensis	Muricauda lutaonensis CC-HSB-11	29	Carotenoid	28
Terpene	4	k141_1769	cut off	s__Sulfitobacter_sp._AM1-D1	Roseobacter sp. AzwK-3b	17		
Terpene	9	k141_2443	no	g__Rhizobium	Mesorhizobium amorphae CCNWGS0123	12		
Terpene	10	k141_3305	no	s__Muricauda_ruestringensis	Imtechella halotolerans K1	17	Carotenoid	28

Terpene	18	k141_6681	no	s__Mesorhizobium_amorphae	Mesorhizobium huakuii 7653R	11		
Terpene	22	k141_9026	end of contig	NA	Roseovarius sp. TM1035	4	Carotenoid	100
Terpene	25	k141_11338	no	s__Roseobacter_denitrificans	Thalassobium sp. R2A62	15		
Terpene	33	k141_16265	cut off at beginning	NA	Flagellimonas sp. DK16	5		
Terpene	37	k141_18837	no	s__Octadecabacter_arcticus	Ruegeria sp. CECT 5091	7		
Terpene	42	k141_22409	no	s__Maribacter_sp._HTCC2170	Flagellimonas sp. DK169	13		
Terpene	44	k141_25046	cut off at beginning	s__Muricauda_lutaonensis	Flagellimonas sp. DK169	29		
Terpene	56	k141_30848	nearly entire contig	s__Bradyrhizobium_sp._ORS_278	Roseovarius tolerans EL-164	4	Carotenoid	100
Terpene	67	k141_36688	no	s__Leisingera_methylohalidivorans	Rhodobacteraceae bacterium KLH11	7		
Terpene	71	k141_39249	no	NA	Roseobacter sp. AzwK-3b	27	Carotenoid	100
Terpene	72	k141_39369	no	s__Roseobacter_litoralis	Confluentimicrobium sp. EMB200-NS6	7		
Terpene	73	k141_40087	no	s__Antarctobacter_heliothermus	Ruegeria sp. ANG-R	41		
Terpene	83	k141_45903	no	s__Roseovarius_mucosus	Roseobacter sp. AzwK-3b	10	Carotenoid	100
Terpene	85	k141_46844	no	s__Phaeobacter_gallaeciensis	Candidatus Rhodobacter lobularis	12		

Terpene	89	k141_49253	no	s__Octadecabacter_arcticus	Silicibacter lacuscaerulensis ITI-1157	6		
Terpene	105	k141_61761	no	s__Marinovum_algicola	Sulfitobacter geojensis MM-124	5		
Terpene	111	k141_68374	end of contig	NA	Erythrobacter litoralis HTCC2594	4		
Terpene	133	k141_86707	no	NA	No significant ClusterBlast hits found			
Terpene	143	k141_94107	beginning of contig	s__Erythrobacter_flavus	Erythrobacter litoralis HTCC2594	16	Ast-axanthin dideoxy-glycoside	75
Terpene	152	k141_99097	end of contig	NA	Muricauda ruestringensis DSM 13258	4	Carotenoid	28
Terpene	161	k141_107066	no	s__Mesorhizobium_amorphae	Nitratireductor indicus C115	38		
Terpene	163	k141_107553	no	g__Maribacter	Zobellia galactanivorans DsiJT	18	Carotenoid	28
Terpene	165	k141_108469	end of contig	s__Ruegeria_mobilis	Ruegeria atlantica CECT4293	7		
Terpene	169	k141_111340	no	s__Gynuella_sunshinyii	Alpha proteobacterium Q-1	28		
Terpene	170	k141_112745	no	s__Ruegeria_pomeroyi	Ruegeria sp. ANG-R	43		
Terpene	174	k141_116241	end of contig	NA	Roseovarius atlanticus R12B	9	Carotenoid	100

Terpene	182	k141_120852	no	s__Tateyamaria_omphalii	Roseovarius sp. 217	34	Carotenoid	100
Terpene	212	k141_150205	beginning of contig	NA	Roseovarius sp. TM1035	4	Carotenoid	100
Terpene	213	k141_150413	beginning of contig	NA	Roseobacter sp. AzwK-3b	5	Carotenoid	100
Terpene	230	k141_163370	end of contig	s__Tateyamaria_omphalii	Tateyamaria sp. ANG-S1	7	Carotenoid	100
Terpene	232	k141_167445	no	s__Muricauda_ruestringensis	Zobellia galactanivorans DsiJT	18	Carotenoid	28
Terpene	238	k141_173409	nearly entire contig	s__Flavisolibacter_tropicus	Flagellimonas sp. DK169	5		
Terpene	241	k141_175588	no	s__Marinovum_algicola	Silicibacter lacuscaerulensis ITI-1157	8		
Terpene	243	k141_177169	no	s__Phaeobacter_gallaeciensis	Silicibacter lacuscaerulensis ITI-1157	6		
Terpene	251	k141_184734	no	s__Ruegeria_pomeroyi	Ruegeria pomeroyi DSS-3	48		
Thiopeptide	86	k141_47797	no	s__Rhizobium_phaseoli	Ruegeria sp. ANG-S4	69		
Thiopeptide	93	k141_52504	no	s__Ruegeria_pomeroyi	Ruegeria sp. ANG-R	28		
Thiopeptide	109	k141_65580	no	s__Leisingera_methylohalidivorans	Leisingera sp. ANG-DT	27		

Appendix IV: Percent of secondary metabolite genes expressed in the ANG and JC metatranscriptomes

Cluster number	gene cluster	Total number of genes in cluster	Number of biosynthetic genes	Percent of total genes expressed				Percent of biosynthetic genes expressed			
				ANG	Day 0 JC	Day 10 control JC	Day 10 challenge JC	ANG	Day 0 JC	Day 10 control JC	Day 10 challenge JC
1	terpene	2	1	50.0	0.0	0.0	50.0	100.0	0.0	0.0	0.0
2	bacteriocin	6	3	16.7	83.3	66.7	83.3	0.0	66.7	33.3	66.7
3	terpene	22	1	4.5	9.1	13.6	27.3	0.0	0.0	0.0	0.0
4	terpene	7	1	85.7	100.0	100.0	100.0	0.0	100.0	100.0	100.0
5	other	5	2	40.0	40.0	80.0	80.0	50.0	50.0	50.0	50.0
6	HSL	5	1	40.0	80.0	80.0	80.0	100.0	100.0	100.0	100.0
7	HSL	14	1	0.0	0.0	28.6	42.9	0.0	0.0	0.0	100.0
8	HSL	20	2	85.0	90.0	85.0	90.0	50.0	50.0	0.0	50.0
9	terpene	7	1	57.1	100.0	100.0	100.0	0.0	100.0	100.0	100.0
10	terpene	7	1	0.0	71.4	85.7	100.0	0.0	100.0	100.0	100.0
11	HSL	7	1	71.4	85.7	71.4	57.1	100.0	100.0	100.0	100.0
12	HSL	22	2	68.2	72.7	77.3	72.7	50.0	100.0	100.0	100.0
13	HSL	26	1	42.3	84.6	84.6	84.6	100.0	100.0	100.0	100.0
14	T3PKS	3	2	66.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0
15	NRPS	15	2	53.3	73.3	80.0	73.3	100.0	100.0	100.0	100.0
16	Bacteriocin	5	1	80.0	100.0	80.0	100.0	100.0	100.0	100.0	100.0
17	HSL	6	1	83.3	50.0	50.0	83.3	100.0	0.0	100.0	100.0
18	terpene	7	1	57.1	71.4	57.1	57.1	0.0	0.0	100.0	100.0
19	Bacteriocin	3	1	66.7	33.3	33.3	66.7	100.0	0.0	0.0	100.0
20	HSL	3	1	66.7	33.3	33.3	66.7	100.0	100.0	100.0	100.0
21	T3PKS-arylpolyene	27	5	7.4	11.1	14.8	22.2	0.0	0.0	0.0	20.0
22	terpene	2	1	0.0	100.0	50.0	50.0	0.0	100.0	100.0	100.0
23	T1PKS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
24	HSL	2	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

25	terpene	25	1	88.0	76.0	68.0	72.0	0.0	0.0	0.0	0.0
26	HSL	9	2	66.7	77.8	55.6	88.9	50.0	50.0	0.0	50.0
27	HSL	4	1	50.0	50.0	50.0	50.0	100.0	100.0	100.0	100.0
28	Bacteriocin	13	3	53.8	38.5	38.5	30.8	0.0	0.0	0.0	33.3
29	ectoine	7	2	57.1	85.7	85.7	85.7	0.0	50.0	50.0	50.0
30	HSL	5	1	60.0	80.0	60.0	60.0	100.0	100.0	100.0	100.0
31	T1PKS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
32	HSL	9	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
33	terpene	3	1	0.0	0.0	33.3	66.7	0.0	0.0	100.0	100.0
34	HSL	9	2	88.9	77.8	88.9	88.9	50.0	0.0	50.0	50.0
35	Bacteriocin	3	1	33.3	0.0	0.0	0.0	100.0	0.0	0.0	0.0
36	Bacteriocin	4	1	75.0	50.0	50.0	50.0	100.0	100.0	100.0	100.0
37	terpene	3	1	66.7	66.7	100.0	100.0	100.0	100.0	100.0	100.0
38	Bacteriocin	15	3	86.7	80.0	66.7	73.3	100.0	66.7	66.7	66.7
39	cyanobactin	2	1	0.0	0.0	50.0	50.0	0.0	0.0	100.0	100.0
40	T1PKS	1	1	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0
41	NRPS	14	2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
42	terpene	10	2	10.0	30.0	10.0	50.0	0.0	0.0	50.0	50.0
43	HSL	2	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
44	terpene	20	4	30.0	20.0	60.0	60.0	0.0	25.0	50.0	25.0
45	HSL	17	2	82.4	70.6	64.7	64.7	100.0	100.0	100.0	100.0
46	ectoine	5	1	60.0	60.0	60.0	60.0	100.0	0.0	100.0	0.0
47	Bacteriocin	2	1	50.0	100.0	100.0	100.0	0.0	100.0	100.0	100.0
48	NRPS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
49	Bacteriocin	2	1	50.0	100.0	0.0	50.0	0.0	100.0	0.0	0.0
50	HSL	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
51	other	54	1	90.7	83.3	79.6	79.6	100.0	100.0	0.0	100.0
52	other	1	1	100.0	100.0	100.0	0.0	100.0	100.0	100.0	0.0
53	HSL	18	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
54	HSL	23	3	65.2	73.9	87.0	91.3	66.7	66.7	66.7	66.7
55	T3PKS	32	5	9.4	21.9	34.4	31.3	0.0	0.0	0.0	0.0

56	terpene	2	1	50.0	100.0	100.0	100.0	0.0	100.0	100.0	100.0
57	HSL	5	1	40.0	80.0	20.0	80.0	100.0	100.0	0.0	100.0
58	Bacteriocin	11	1	63.6	18.2	36.4	81.8	0.0	0.0	0.0	100.0
59	Bacteriocin	3	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
60	HSL	10	2	90.0	80.0	90.0	80.0	50.0	50.0	100.0	50.0
61	other	5	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
62	HSL	19	2	84.2	84.2	57.9	73.7	50.0	50.0	50.0	50.0
63	Bacteriocin	10	1	70.0	60.0	70.0	60.0	0.0	0.0	100.0	100.0
64	Bacteriocin	7	1	57.1	85.7	85.7	100.0	0.0	0.0	100.0	100.0
65	HSL	2	1	100.0	100.0	0.0	100.0	100.0	100.0	0.0	100.0
66	Bacteriocin	14	3	35.7	64.3	57.1	64.3	66.7	66.7	66.7	66.7
67	terpene	3	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
68	HSL	16	2	68.8	75.0	87.5	87.5	100.0	100.0	100.0	100.0
69	other	4	1	75.0	75.0	50.0	75.0	100.0	100.0	100.0	100.0
70	arylpolyene	38	2	63.2	73.7	73.7	81.6	100.0	100.0	100.0	100.0
71	terpene	11	1	27.3	18.2	36.4	45.5	0.0	0.0	0.0	100.0
72	terpene	12	1	75.0	75.0	75.0	75.0	0.0	100.0	0.0	0.0
73	terpene	17	2	88.2	88.2	76.5	94.1	100.0	100.0	50.0	100.0
74	nucleoside	3	1	100.0	33.3	33.3	66.7	100.0	0.0	0.0	0.0
75	Bacteriocin	6	1	83.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0
76	Bacteriocin	4	1	0.0	0.0	50.0	75.0	0.0	0.0	100.0	0.0
77	NRPS	15	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
78	Bacteriocin	7	1	0.0	0.0	0.0	28.6	0.0	0.0	0.0	0.0
79	siderophore	6	2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
80	T1PKS	3	1	0.0	0.0	0.0	33.3	0.0	0.0	0.0	0.0
81	HSL	8	1	25.0	25.0	37.5	50.0	0.0	0.0	0.0	0.0
82	Bacteriocin	6	1	83.3	83.3	100.0	100.0	100.0	100.0	100.0	100.0
83	terpene	4	1	0.0	25.0	25.0	25.0	0.0	0.0	0.0	0.0
84	arylpolyene	4	2	0.0	0.0	25.0	50.0	0.0	0.0	0.0	50.0
85	terpene	14	1	71.4	71.4	64.3	64.3	100.0	0.0	0.0	0.0
86	thiopeptide	35	12	74.3	82.9	68.6	71.4	41.7	66.7	66.7	75.0

87	arylpolyene	9	4	0.0	55.6	55.6	66.7	0.0	75.0	75.0	100.0
88	Bacteriocin	3	1	100.0	33.3	100.0	100.0	100.0	0.0	100.0	100.0
89	terpene	3	1	0.0	66.7	66.7	66.7	0.0	100.0	100.0	100.0
90	NRPS	13	2	53.8	92.3	76.9	76.9	100.0	100.0	50.0	100.0
91	HSL	5	2	40.0	80.0	80.0	60.0	0.0	50.0	50.0	50.0
92	HSL	29	2	31.0	75.9	69.0	69.0	100.0	100.0	100.0	100.0
93	thiopeptide	23	3	82.6	82.6	82.6	87.0	100.0	100.0	100.0	100.0
94	Bacteriocin	4	1	75.0	75.0	75.0	75.0	100.0	100.0	100.0	100.0
95	Bacteriocin	2	1	0.0	50.0	0.0	0.0	0.0	0.0	0.0	0.0
96	T1PKS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
97	T3PKS	3	1	66.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0
98	arylpolyene	6	3	66.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0
99	T1PKS	23	2	87.0	69.6	69.6	69.6	100.0	100.0	100.0	50.0
100	other	4	1	50.0	0.0	25.0	0.0	100.0	0.0	0.0	0.0
101	other	12	2	83.3	58.3	41.7	50.0	100.0	50.0	0.0	0.0
102	other	8	1	100.0	62.5	75.0	50.0	100.0	0.0	0.0	0.0
103	other	4	1	75.0	75.0	75.0	75.0	100.0	100.0	100.0	100.0
104	arylpolyene	40	8	70.0	15.0	50.0	95.0	100.0	25.0	87.5	100.0
105	terpene	6	1	50.0	66.7	100.0	100.0	0.0	0.0	100.0	100.0
106	NRPS	1	1	100.0	0.0	0.0	100.0	100.0	0.0	0.0	100.0
107	HSL	4	1	75.0	100.0	100.0	100.0	0.0	100.0	100.0	100.0
108	other	25	1	36.0	52.0	44.0	48.0	0.0	0.0	0.0	0.0
109	thiopeptide	29	2	86.2	82.8	79.3	75.9	100.0	100.0	100.0	100.0
110	other	3	2	66.7	33.3	33.3	66.7	100.0	50.0	50.0	100.0
111	terpene	2	1	0.0	0.0	0.0	50.0	0.0	0.0	0.0	0.0
112	NRPS	46	4	58.7	52.2	47.8	47.8	25.0	50.0	25.0	50.0
113	NRPS	8	3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
114	Bacteriocin	3	1	100.0	66.7	100.0	66.7	100.0	0.0	100.0	0.0
115	T3PKS	2	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
116	Bacteriocin	8	1	87.5	75.0	37.5	75.0	100.0	100.0	0.0	100.0
117	Bacteriocin	2	1	100.0	100.0	50.0	0.0	100.0	100.0	100.0	0.0

118	NRPS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
119	other	3	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
120	T1PKS	26	8	88.5	88.5	76.9	88.5	100.0	100.0	100.0	100.0
121	ectoine	10	3	60.0	50.0	40.0	60.0	33.3	33.3	0.0	33.3
122	HSL	5	2	60.0	40.0	80.0	80.0	50.0	0.0	100.0	100.0
123	ectoine	13	1	7.7	0.0	46.2	46.2	0.0	0.0	0.0	100.0
124	other	5	1	80.0	80.0	60.0	60.0	100.0	100.0	100.0	100.0
125	NRPS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
126	other	15	5	60.0	40.0	66.7	73.3	60.0	40.0	60.0	60.0
127	arylpolyene	6	2	66.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0
128	NRPS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
129	HSL	2	1	100.0	100.0	50.0	100.0	100.0	100.0	100.0	100.0
130	NRPS	12	3	33.3	75.0	83.3	66.7	33.3	66.7	66.7	66.7
131	HSL	11	1	18.2	18.2	45.5	27.3	0.0	0.0	0.0	0.0
132	NRPS	2	2	50.0	0.0	0.0	100.0	50.0	0.0	0.0	100.0
133	terpene	8	3	12.5	37.5	25.0	25.0	0.0	33.3	33.3	33.3
134	HSL	3	1	100.0	33.3	66.7	66.7	100.0	0.0	0.0	100.0
135	HSL	19	2	52.6	47.4	47.4	47.4	100.0	100.0	50.0	50.0
136	Bacteriocin	2	1	50.0	0.0	0.0	50.0	0.0	0.0	0.0	100.0
137	T1PKS	1	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
138	HSL	2	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
139	HSL	10	2	90.0	80.0	80.0	90.0	100.0	50.0	100.0	100.0
140	Bacteriocin	2	1	0.0	100.0	100.0	100.0	0.0	100.0	100.0	100.0
141	HSL	12	1	75.0	66.7	50.0	50.0	100.0	0.0	0.0	0.0
142	other	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
143	terpene	8	2	37.5	12.5	37.5	25.0	0.0	0.0	0.0	0.0
144	HSL	17	1	94.1	94.1	94.1	100.0	100.0	100.0	100.0	100.0
145	HSL	7	1	14.3	42.9	42.9	57.1	0.0	0.0	0.0	0.0
146	ectoine	13	1	38.5	23.1	30.8	46.2	100.0	100.0	0.0	100.0
147	other	3	1	66.7	0.0	100.0	0.0	100.0	0.0	100.0	0.0
148	Bacteriocin	6	3	0.0	16.7	50.0	83.3	0.0	33.3	66.7	100.0

149	Bacteriocin	10	1	80.0	70.0	70.0	70.0	0.0	0.0	0.0	0.0
150	Bacteriocin	3	1	66.7	66.7	33.3	66.7	0.0	0.0	0.0	0.0
151	HSL	2	1	100.0	100.0	50.0	100.0	100.0	100.0	0.0	100.0
152	terpene	2	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
153	T1PKS-NRPS	6	2	50.0	100.0	100.0	100.0	50.0	100.0	100.0	100.0
154	siderophore	1	1	100.0	0.0	100.0	100.0	100.0	0.0	100.0	100.0
155	Bacteriocin	3	1	33.3	0.0	66.7	100.0	100.0	0.0	100.0	100.0
156	NRPS	25	3	96.0	92.0	96.0	92.0	100.0	100.0	100.0	100.0
157	T1PKS	11	4	72.7	81.8	45.5	72.7	75.0	50.0	50.0	50.0
158	Bacteriocin	16	2	81.3	75.0	81.3	87.5	100.0	100.0	100.0	100.0
159	Bacteriocin	17	2	29.4	52.9	58.8	70.6	0.0	50.0	50.0	100.0
160	NRPS	7	1	57.1	14.3	57.1	100.0	100.0	0.0	0.0	100.0
161	terpene	22	1	63.6	86.4	68.2	86.4	100.0	100.0	0.0	100.0
162	Bacteriocin	2	1	100.0	50.0	50.0	50.0	100.0	100.0	0.0	100.0
163	terpene	8	1	25.0	62.5	75.0	75.0	0.0	100.0	100.0	100.0
164	ectoine	14	1	35.7	42.9	50.0	50.0	100.0	0.0	100.0	100.0
165	terpene	3	1	66.7	66.7	66.7	100.0	0.0	0.0	0.0	100.0
166	siderophore	2	2	0.0	0.0	50.0	50.0	0.0	0.0	50.0	50.0
167	other	21	7	42.9	47.6	28.6	47.6	57.1	71.4	57.1	71.4
168	HSL	4	1	100.0	100.0	50.0	100.0	100.0	100.0	0.0	100.0
169	terpene	16	1	18.8	25.0	31.3	31.3	0.0	0.0	0.0	0.0
170	terpene	23	2	87.0	91.3	82.6	82.6	100.0	100.0	100.0	100.0
171	HSL	18	4	94.4	94.4	94.4	94.4	100.0	100.0	100.0	100.0
172	NRPS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
173	siderophore	2	1	50.0	50.0	0.0	0.0	100.0	100.0	0.0	0.0
174	terpene	3	1	33.3	33.3	33.3	66.7	0.0	0.0	0.0	0.0
175	Bacteriocin	7	1	57.1	85.7	42.9	100.0	0.0	100.0	0.0	100.0
176	Bacteriocin	7	1	85.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0
177	HSL	22	1	4.5	18.2	9.1	9.1	0.0	0.0	0.0	0.0
178	HSL	16	1	62.5	81.3	81.3	93.8	100.0	100.0	0.0	100.0
179	arylpolyene	4	3	0.0	25.0	50.0	75.0	0.0	0.0	66.7	66.7

180	acyl_amino_acids	9	2	88.9	100.0	88.9	88.9	100.0	100.0	100.0	100.0
181	HSL	4	1	75.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
182	terpene	20	2	25.0	65.0	80.0	80.0	50.0	100.0	50.0	100.0
183	Bacteriocin	15	3	53.3	46.7	66.7	60.0	66.7	33.3	66.7	66.7
184	ectoine	10	2	30.0	40.0	60.0	60.0	0.0	0.0	50.0	50.0
185	NRPS	3	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
186	other	2	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
187	HSL	2	1	0.0	100.0	100.0	100.0	0.0	100.0	100.0	100.0
188	Bacteriocin	6	1	33.3	33.3	33.3	83.3	0.0	0.0	0.0	100.0
189	other	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
190	HSL	14	1	92.9	92.9	85.7	92.9	100.0	100.0	100.0	100.0
191	siderophore	3	2	100.0	100.0	66.7	66.7	100.0	100.0	100.0	100.0
192	HSL	18	1	5.6	16.7	16.7	22.2	0.0	0.0	0.0	0.0
193	other	41	1	97.6	82.9	70.7	65.9	100.0	100.0	100.0	100.0
194	NRPS	27	4	55.6	51.9	48.1	48.1	100.0	100.0	75.0	75.0
195	siderophore	4	3	75.0	50.0	50.0	50.0	66.7	33.3	33.3	33.3
196	other	13	2	30.8	38.5	46.2	46.2	0.0	50.0	50.0	50.0
197	T1PKS-NRPS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
198	HSL	12	2	91.7	66.7	75.0	75.0	50.0	0.0	50.0	50.0
199	ectoine	13	1	53.8	38.5	46.2	61.5	0.0	100.0	100.0	100.0
200	other	2	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
201	T1PKS	1	1	0.0	100.0	100.0	100.0	0.0	100.0	100.0	100.0
202	HSL	7	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
203	HSL	13	2	76.9	76.9	69.2	84.6	100.0	50.0	50.0	50.0
204	ectoine	5	1	80.0	80.0	60.0	80.0	0.0	0.0	0.0	100.0
205	HSL	2	1	100.0	0.0	100.0	50.0	100.0	0.0	100.0	0.0
206	arylpolyene	4	2	0.0	50.0	75.0	75.0	0.0	100.0	100.0	100.0
207	other	10	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
208	HSL	25	2	72.0	88.0	92.0	96.0	50.0	100.0	100.0	100.0
209	T1PKS	2	1	50.0	50.0	50.0	50.0	100.0	100.0	100.0	100.0

210	siderophore	3	2	100.0	66.7	66.7	66.7	100.0	100.0	100.0	100.0
211	arylpolyene	5	3	100.0	100.0	100.0	80.0	100.0	100.0	100.0	100.0
212	terpene	2	1	100.0	50.0	0.0	50.0	100.0	100.0	0.0	100.0
213	terpene	2	1	50.0	0.0	100.0	50.0	100.0	0.0	100.0	100.0
214	HSL	2	1	50.0	50.0	50.0	50.0	0.0	0.0	100.0	0.0
215	T1PKS	2	1	50.0	50.0	50.0	50.0	0.0	0.0	0.0	0.0
216	Bacteriocin	2	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
217	NRPS	42	4	69.0	92.9	90.5	95.2	50.0	75.0	75.0	75.0
218	HSL	17	1	47.1	35.3	41.2	29.4	100.0	0.0	0.0	0.0
219	HSL	5	1	100.0	100.0	80.0	80.0	100.0	100.0	100.0	100.0
220	NRPS	36	13	88.9	91.7	88.9	88.9	92.3	100.0	92.3	92.3
221	ectoine	6	2	16.7	33.3	50.0	66.7	0.0	50.0	100.0	100.0
222	HSL	5	1	40.0	100.0	60.0	60.0	100.0	100.0	0.0	0.0
223	NRPS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
224	NRPS	4	2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
225	HSL	3	1	100.0	100.0	100.0	66.7	100.0	100.0	100.0	100.0
226	other	7	2	28.6	85.7	57.1	85.7	50.0	100.0	100.0	100.0
227	HSL	2	1	50.0	50.0	100.0	100.0	0.0	100.0	100.0	100.0
228	ectoine	3	1	0.0	33.3	0.0	0.0	0.0	0.0	0.0	0.0
229	NRPS	27	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
230	terpene	2	1	100.0	0.0	50.0	50.0	100.0	0.0	0.0	0.0
231	HSL	2	1	50.0	50.0	100.0	100.0	0.0	0.0	100.0	100.0
232	terpene	7	1	42.9	85.7	85.7	85.7	0.0	100.0	100.0	100.0
233	HSL	15	2	13.3	26.7	46.7	60.0	0.0	50.0	100.0	50.0
234	HSL	3	1	33.3	66.7	66.7	66.7	100.0	100.0	100.0	100.0
235	Bacteriocin	4	1	50.0	50.0	50.0	100.0	100.0	0.0	100.0	100.0
236	Bacteriocin	11	3	90.9	72.7	54.5	45.5	100.0	66.7	33.3	66.7
237	HSL	14	2	0.0	7.1	7.1	7.1	0.0	0.0	0.0	0.0
238	terpene	2	1	0.0	50.0	50.0	50.0	0.0	100.0	100.0	100.0
239	T1PKS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
240	NRPS	4	2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

241	terpene	4	1	75.0	100.0	100.0	100.0	0.0	100.0	100.0	100.0
242	other	2	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
243	terpene	3	1	100.0	33.3	100.0	33.3	100.0	0.0	100.0	0.0
244	HSL	22	1	4.5	0.0	9.1	13.6	0.0	0.0	0.0	0.0
245	NRPS	17	2	0.0	0.0	0.0	5.9	0.0	0.0	0.0	50.0
246	T1PKS	21	4	57.1	66.7	76.2	90.5	25.0	25.0	75.0	75.0
247	T1PKS	1	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
248	HSL	4	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
249	Bacteriocin	2	1	50.0	50.0	50.0	50.0	0.0	0.0	0.0	0.0
250	other	9	5	100.0	77.8	77.8	77.8	100.0	80.0	80.0	60.0
251	terpene	26	2	50.0	69.2	84.6	88.5	50.0	100.0	100.0	100.0
252	Bacteriocin	12	2	66.7	75.0	83.3	83.3	100.0	50.0	100.0	100.0
253	Bacteriocin	14	2	50.0	71.4	78.6	78.6	50.0	50.0	50.0	50.0
254	Bacteriocin	12	1	25.0	50.0	66.7	75.0	0.0	100.0	100.0	0.0
255	HSL	3	1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Appendix V: *Verrucomicrobia* PD13 Bin 1 MAG gene expression (average TPM values) in ANG and JC metatranscriptomes. Values in table correspond to heatmaps in Chapter 4, Figures 10-12. Data is organized alphabetically by heatmap title and row number.

Heatmap name	Row #	Gene	ANG	ANG st dev	Day 0 JC	D0 JC st dev	D10 Con JC	D10 Con JC st dev	D10 Chall JC	D10 Chall JC st dev
Arylsulfatase	1	Arylsulfatase	39.0	24.6	33.8	10.4	51.8	32.6	47.1	26.5
Arylsulfatase	2	Arylsulfatase	56.2	36.0	54.5	35.3	61.4	24.2	66.7	26.5
Arylsulfatase	3	Arylsulfatase	34.4	22.8	37.7	33.6	48.3	29.4	49.2	20.7
Arylsulfatase	4	Arylsulfatase	30.1	21.4	22.0	15.3	18.0	15.0	21.8	18.4
Arylsulfatase	5	Arylsulfatase	37.8	30.7	30.3	16.5	24.6	9.0	16.9	7.6
Arylsulfatase	6	Arylsulfatase	37.0	15.1	32.0	11.3	30.4	14.7	25.9	18.3
Arylsulfatase	7	Arylsulfatase	38.5	23.9	27.3	18.4	47.4	25.7	48.4	29.1
Arylsulfatase	8	Arylsulfatase	45.8	17.3	28.5	15.5	47.4	18.8	52.7	20.5
Arylsulfatase	9	Arylsulfatase	52.6	43.1	34.1	36.5	29.3	29.3	30.0	27.0
Arylsulfatase	10	Arylsulfatase	31.3	13.6	28.8	13.1	26.6	11.5	32.9	12.7
Arylsulfatase	11	Arylsulfatase	13.9	11.6	12.2	12.2	13.0	12.9	18.6	17.1
Arylsulfatase	12	Arylsulfatase	40.7	29.6	27.7	25.7	42.5	48.3	36.6	38.1
Arylsulfatase	13	Arylsulfatase	30.7	17.2	27.4	26.1	36.2	17.9	41.3	12.4
Arylsulfatase	14	Arylsulfatase	42.9	30.5	24.9	16.0	30.8	14.1	30.0	15.9
Arylsulfatase	15	Arylsulfatase	25.5	14.9	24.3	10.6	28.0	10.6	22.6	14.7
Arylsulfatase	16	Arylsulfatase	14.1	7.0	10.2	5.8	13.3	8.3	13.7	5.2
Arylsulfatase	17	Arylsulfatase	7.8	6.1	4.3	5.7	6.1	7.9	16.3	17.8
Arylsulfatase	18	Arylsulfatase	6.8	6.1	5.8	3.4	11.0	4.8	13.0	7.3
Arylsulfatase	19	Arylsulfatase	43.7	34.2	34.0	20.3	21.8	12.8	21.8	12.2
Arylsulfatase	20	Arylsulfatase	80.9	75.0	91.4	63.4	55.4	47.2	57.5	50.7
Arylsulfatase	21	Arylsulfatase	31.6	14.6	25.1	15.1	23.8	15.5	30.6	14.7

Arylsulfatase	22	Arylsulfatase	16.3	13.9	14.1	13.0	19.4	12.2	23.1	16.3
Arylsulfatase	23	Arylsulfatase	25.0	17.2	20.8	15.0	24.2	10.6	24.7	10.6
Arylsulfatase	24	Arylsulfatase	43.2	28.5	30.1	15.9	82.5	44.6	81.8	43.7
Arylsulfatase	25	Arylsulfatase	11.3	7.1	13.5	6.7	17.6	12.0	12.7	2.9
Arylsulfatase	26	Arylsulfatase	8.0	5.7	8.2	6.5	11.8	8.1	11.8	6.3
Arylsulfatase	27	Arylsulfatase	18.6	9.8	31.7	26.0	32.6	15.1	34.1	10.2
Arylsulfatase	28	Arylsulfatase	9.7	4.3	9.4	7.4	15.2	7.1	19.3	12.4
Arylsulfatase	29	Arylsulfatase	13.2	7.6	14.9	9.7	17.5	11.1	19.4	13.1
Arylsulfatase	30	Arylsulfatase	33.5	15.2	34.0	17.7	35.5	17.6	37.4	11.7
Arylsulfatase	31	Arylsulfatase	17.6	11.0	19.7	10.6	18.5	11.3	27.7	9.7
Arylsulfatase	32	Arylsulfatase	68.7	48.3	51.2	15.8	36.8	16.4	33.8	14.3
Arylsulfatase	33	Arylsulfatase	32.0	15.7	25.9	16.8	26.1	9.1	39.0	20.4
Arylsulfatase	34	Arylsulfatase	22.6	17.7	11.1	8.2	11.0	9.5	7.1	3.7
Arylsulfatase	35	Arylsulfatase	392.8	203.1	400.0	333.0	709.0	391.4	723.5	394.2
Arylsulfatase	36	Arylsulfatase	65.8	44.1	64.4	42.6	59.1	41.7	64.9	49.1
Arylsulfatase	37	Arylsulfatase	12.9	10.3	9.1	7.6	10.5	7.3	8.5	5.9
Arylsulfatase	38	Arylsulfatase	32.0	13.1	26.4	16.3	28.4	12.8	33.4	18.3
Arylsulfatase	39	Arylsulfatase	25.1	11.1	20.2	14.6	27.3	13.9	24.3	11.9
Arylsulfatase	40	Arylsulfatase	70.1	42.7	93.1	51.7	55.1	22.7	54.5	24.9
Arylsulfatase	41	Arylsulfatase	19.7	10.3	17.7	12.2	13.2	7.1	14.7	3.9
Arylsulfatase	42	Arylsulfatase	39.9	30.2	41.3	22.7	28.7	20.4	32.0	26.6
Arylsulfatase	43	Arylsulfatase	26.4	18.4	27.2	25.1	30.2	19.3	24.4	18.1
Arylsulfatase	44	Arylsulfatase	28.3	14.4	24.7	17.3	24.0	11.7	27.6	13.8
Arylsulfatase	45	Arylsulfatase	6.5	5.2	7.5	5.5	8.7	7.7	10.4	13.3
Arylsulfatase	46	Arylsulfatase	93.4	43.3	113.0	82.7	77.0	47.6	75.8	38.3
Arylsulfatase	47	Arylsulfatase	28.9	14.4	23.6	13.1	25.2	15.1	38.6	22.2

Arylsulfatase	48	Arylsulfatase	23.9	12.7	24.1	8.2	20.4	11.3	21.3	9.0
Arylsulfatase	49	Arylsulfatase	9.6	6.0	10.8	8.2	8.7	7.2	16.8	19.5
Arylsulfatase	50	Arylsulfatase	25.8	14.6	36.3	34.3	45.0	41.2	44.1	30.6
Arylsulfatase	51	Arylsulfatase	10.3	4.7	8.3	6.1	13.3	10.5	12.4	8.6
Arylsulfatase	52	Arylsulfatase	22.1	9.6	21.6	14.2	26.3	14.6	32.0	18.3
Arylsulfatase	53	Arylsulfatase	26.7	20.2	21.6	9.4	20.7	9.7	22.2	12.3
Arylsulfatase	54	Arylsulfatase	21.3	18.6	8.2	9.6	15.5	14.8	16.0	16.3
Arylsulfatase	55	Arylsulfatase	12.1	12.3	12.4	14.8	27.0	37.7	18.2	19.7
Arylsulfatase	56	Arylsulfatase	29.8	18.8	25.9	15.8	28.0	16.8	33.6	18.1
Arylsulfatase	57	Arylsulfatase	56.3	24.1	57.6	36.0	46.3	24.7	58.0	31.8
Arylsulfatase	58	Arylsulfatase	33.1	23.2	22.4	21.0	30.3	17.0	32.3	18.1
Arylsulfatase	59	Arylsulfatase	70.8	59.3	59.2	41.5	96.7	68.2	97.3	70.5
Arylsulfatase	60	Arylsulfatase	130.2	56.4	112.5	74.7	254.9	107.7	280.2	144.2
Arylsulfatase	61	Arylsulfatase	29.2	14.2	22.0	13.9	44.3	23.8	46.3	20.7
Arylsulfatase	62	Arylsulfatase	32.6	16.7	25.6	15.6	36.5	15.8	40.0	12.1
Arylsulfatase	63	Arylsulfatase	17.7	7.5	16.8	7.8	26.0	13.9	24.2	8.6
Arylsulfatase	64	Arylsulfatase	17.1	5.8	15.9	8.6	13.5	6.0	19.9	5.6
Arylsulfatase	65	Arylsulfatase	34.4	8.8	28.3	6.1	27.2	12.1	29.0	6.4
Arylsulfatase	66	Arylsulfatase	38.6	6.9	27.6	13.1	30.0	11.8	30.5	11.4
Arylsulfatase	67	Arylsulfatase	29.7	16.2	18.5	13.6	21.6	14.3	20.9	10.5
Arylsulfatase	68	Arylsulfatase	29.0	9.9	20.7	11.6	22.4	11.4	24.3	6.8
Arylsulfatase	69	Arylsulfatase	60.3	31.8	60.4	35.4	40.0	23.1	55.5	35.5
Arylsulfatase	70	Arylsulfatase	33.5	19.1	25.9	19.1	37.7	20.3	34.2	19.9
Arylsulfatase	71	Arylsulfatase	45.8	24.6	33.5	24.4	36.4	19.3	38.2	18.7
Arylsulfatase	72	Arylsulfatase	15.8	9.2	12.6	5.6	7.9	2.8	14.0	5.0
Arylsulfatase	73	Arylsulfatase	24.0	16.1	24.6	15.1	37.4	26.9	35.9	18.8

Arylsulfatase	74	Arylsulfatase	497.8	374.0	909.8	609.0	761.4	352.5	848.1	282.8
Arylsulfatase	75	Arylsulfatase	84.5	54.6	78.9	63.7	99.1	62.3	102.0	45.3
Arylsulfatase	76	Arylsulfatase	18.2	13.2	19.0	16.3	38.1	23.8	31.9	26.8
Arylsulfatase	77	Arylsulfatase	40.6	24.0	35.7	30.3	44.0	23.1	54.5	26.5
Arylsulfatase	78	Arylsulfatase	15.2	12.2	13.3	5.7	16.8	8.9	15.0	4.2
Arylsulfatase	79	Arylsulfatase	43.1	17.8	33.7	17.5	31.4	12.0	29.0	13.2
Arylsulfatase	80	Arylsulfatase	12.7	9.3	13.1	7.4	12.5	9.8	14.6	15.7
Arylsulfatase	81	Arylsulfatase	55.1	40.6	45.2	45.5	63.9	61.6	59.3	56.1
Arylsulfatase	82	Arylsulfatase	17.6	17.9	9.5	7.6	11.0	9.9	9.8	8.7
Arylsulfatase	83	Arylsulfatase	22.3	16.6	19.9	16.3	18.3	16.8	21.1	15.9
Arylsulfatase	84	Arylsulfatase	24.6	20.8	13.9	7.9	8.5	5.5	12.2	10.4
Arylsulfatase	85	Arylsulfatase	11.8	7.1	9.9	7.5	14.4	6.5	15.3	2.4
Arylsulfatase	86	Arylsulfatase	13.4	14.0	12.8	9.5	15.5	13.2	18.7	14.4
Arylsulfatase	87	Arylsulfatase	30.6	19.9	27.3	23.0	24.3	16.6	21.7	13.8
Arylsulfatase	88	Arylsulfatase	18.2	9.6	9.8	6.3	9.5	5.0	9.9	5.2
Arylsulfatase	89	Arylsulfatase	16.1	7.0	13.0	3.9	11.6	5.0	10.3	8.3
Arylsulfatase	90	Arylsulfatase	32.6	19.5	32.8	24.2	34.3	12.4	35.8	23.5
Arylsulfatase	91	Arylsulfatase	68.5	44.6	20.8	7.5	19.6	13.3	26.2	18.8
Arylsulfatase	92	Arylsulfatase	18.9	10.7	16.1	6.3	25.0	19.6	25.8	18.0
Arylsulfatase	93	Arylsulfatase	38.5	30.5	24.5	25.0	67.7	59.5	90.1	81.3
Arylsulfatase	94	Arylsulfatase	31.3	13.2	19.2	8.6	20.5	7.8	23.0	4.9
Arylsulfatase	95	Arylsulfatase	15.9	6.0	14.8	10.6	18.0	9.8	16.4	8.1
Arylsulfatase	96	Arylsulfatase	43.0	17.9	36.8	26.7	37.4	14.2	33.9	10.5
Arylsulfatase	97	Arylsulfatase	101.0	48.4	113.7	60.6	91.7	57.2	85.5	50.3
Arylsulfatase	98	Arylsulfatase	44.8	21.6	41.1	26.7	50.9	30.6	44.5	18.7
Arylsulfatase	99	Arylsulfatase	19.4	12.4	20.2	20.3	19.0	14.6	19.5	13.0

Arylsulfatase	100	Arylsulfatase	12.3	8.4	8.8	6.5	4.8	2.7	10.2	12.7
Arylsulfatase	101	Arylsulfatase	9.6	7.4	5.6	4.5	6.9	5.4	7.8	7.4
Arylsulfatase	102	Arylsulfatase	12.0	7.4	5.8	4.2	7.6	7.5	8.6	8.8
Arylsulfatase	103	Arylsulfatase	32.5	23.6	20.6	15.9	19.4	19.5	23.1	18.3
Arylsulfatase	104	Arylsulfatase	20.4	15.3	21.1	17.9	23.4	14.1	22.6	10.4
Arylsulfatase	105	Arylsulfatase	20.0	13.6	20.6	9.1	18.0	9.7	17.2	10.3
Arylsulfatase	106	Arylsulfatase	37.9	17.3	40.5	20.6	50.7	15.0	53.1	14.4
Arylsulfatase	107	Arylsulfatase	35.4	16.0	38.7	24.5	53.4	21.8	59.3	23.7
Arylsulfatase	108	Arylsulfatase	184.4	111.6	156.3	117.0	135.8	99.7	154.1	95.9
Arylsulfatase	109	Arylsulfatase	14.3	9.5	13.3	9.3	24.4	11.1	21.8	8.3
Arylsulfatase	110	Arylsulfatase	38.7	13.8	30.2	17.5	33.7	16.8	35.5	12.9
Arylsulfatase	111	Arylsulfatase	29.9	17.2	22.4	16.1	24.6	11.9	27.1	13.2
Arylsulfatase	112	Arylsulfatase	62.4	21.7	49.8	30.2	63.8	39.5	44.0	22.9
Arylsulfatase	113	Arylsulfatase	34.2	11.1	24.8	17.9	23.9	10.4	22.9	8.4
Arylsulfatase	114	Arylsulfatase	60.1	29.7	53.7	38.6	76.6	52.2	84.1	51.1
Arylsulfatase	115	Arylsulfatase	34.2	22.7	28.1	19.1	32.3	15.8	36.8	19.3
Arylsulfatase	116	Arylsulfatase	7.9	5.8	4.1	3.2	6.7	6.6	4.9	4.5
Arylsulfatase	117	Arylsulfatase	31.0	18.1	27.4	22.4	26.2	11.1	31.1	20.0
Arylsulfatase	118	Arylsulfatase	35.9	35.7	28.4	29.5	15.2	11.7	34.2	34.5
Arylsulfatase	119	Arylsulfatase	108.0	58.9	94.4	66.8	118.5	53.8	116.3	40.8
Arylsulfatase	120	Arylsulfatase	95.5	80.8	118.4	143.4	123.4	124.7	126.8	121.9
Arylsulfatase	121	Arylsulfatase	48.0	42.7	54.0	67.0	64.1	54.6	59.3	56.8
Arylsulfatase	122	Arylsulfatase	42.6	25.8	38.9	42.5	47.2	42.6	48.6	44.9
Arylsulfatase	123	Arylsulfatase	23.1	14.2	24.7	23.8	41.7	36.5	33.6	23.8
Arylsulfatase	124	Arylsulfatase	108.8	136.8	89.6	71.5	58.6	50.2	62.5	51.6
Arylsulfatase	125	Arylsulfatase	20.0	14.2	15.8	14.0	19.0	15.1	19.7	10.6

Arylsulfatase	126	Arylsulfatase	5.7	8.0	5.1	5.8	8.7	8.2	5.3	3.1
Arylsulfatase	127	Arylsulfatase	15.2	10.9	11.5	5.8	14.0	13.2	13.7	11.0
Arylsulfatase	128	Arylsulfatase	28.0	15.4	22.4	13.1	28.5	18.9	27.1	12.3
Arylsulfatase	129	Arylsulfatase	9.4	8.1	8.6	8.4	5.4	5.1	8.5	8.5
Arylsulfatase	130	Arylsulfatase	62.4	45.2	45.9	36.2	56.7	36.8	69.7	27.9
Arylsulfatase	131	Arylsulfatase	9.2	5.0	7.9	6.9	15.5	9.7	14.7	7.2
Arylsulfatase	132	Arylsulfatase	12.0	3.0	11.0	6.1	13.5	7.0	17.7	9.5
Arylsulfatase	133	Arylsulfatase	25.9	13.5	26.4	15.6	46.4	27.6	34.2	16.4
Arylsulfatase	134	Arylsulfatase	44.9	21.5	49.3	37.3	44.8	26.3	56.6	23.7
Arylsulfatase	135	Arylsulfatase	9.8	5.7	13.2	10.6	16.7	8.0	20.7	4.5
Arylsulfatase	136	Arylsulfatase	26.7	11.1	24.8	13.6	24.6	12.1	29.8	10.3
Arylsulfatase	137	Arylsulfatase	18.0	10.6	12.4	9.7	18.4	10.7	15.0	5.0
Arylsulfatase	138	Arylsulfatase	12.8	7.9	9.2	5.9	16.0	10.0	14.8	9.9
Arylsulfatase	139	Arylsulfatase	28.6	17.7	24.6	16.6	25.4	9.6	23.5	8.4
Arylsulfatase	140	Arylsulfatase	62.6	33.4	50.8	17.9	55.9	25.7	55.7	23.6
Arylsulfatase	141	Arylsulfatase	40.6	27.4	33.8	25.9	53.8	39.7	62.7	24.7
Arylsulfatase	142	Arylsulfatase	12.6	8.7	10.8	7.4	11.6	10.4	10.2	9.7
Arylsulfatase	143	Arylsulfatase	9.5	5.3	8.2	4.5	17.8	9.9	11.2	6.9
Arylsulfatase	144	Arylsulfatase	18.1	15.3	11.5	6.5	21.1	10.3	18.3	7.8
Arylsulfatase	145	Arylsulfatase	14.9	11.8	7.6	6.9	15.3	7.3	14.9	5.5
Arylsulfatase	146	Arylsulfatase	15.2	12.1	7.8	9.7	20.9	8.8	15.9	11.1
Arylsulfatase	147	Arylsulfatase	37.7	8.0	29.5	9.4	31.5	12.5	27.4	10.7
Arylsulfatase	148	Arylsulfatase	23.8	15.8	24.5	16.4	33.0	20.8	26.8	9.5
Arylsulfatase	149	Arylsulfatase	30.2	10.7	20.7	8.3	26.5	11.1	26.0	11.9
Arylsulfatase	150	Arylsulfatase	42.7	22.0	37.7	19.3	43.3	17.0	51.1	21.8
Arylsulfatase	151	Arylsulfatase	23.5	14.6	27.5	20.0	39.4	26.8	40.2	24.1

Arylsulfatase	152	Arylsulfatase	54.4	24.8	39.1	8.4	35.5	15.8	44.1	14.0
Arylsulfatase	153	Arylsulfatase	9.0	7.0	8.9	7.4	5.8	3.5	6.7	4.0
Arylsulfatase	154	Arylsulfatase	28.1	11.5	27.8	21.6	20.7	9.0	22.7	14.5
Arylsulfatase	155	Arylsulfatase	21.6	17.4	13.7	11.1	18.5	14.4	14.9	11.8
Arylsulfatase	156	Arylsulfatase	20.6	13.3	15.4	13.3	19.5	12.5	21.7	14.4
Arylsulfatase	157	Arylsulfatase	21.7	8.7	23.9	11.2	18.7	5.2	20.4	7.8
Arylsulfatase	158	Arylsulfatase	31.1	11.3	19.2	7.7	19.3	9.5	18.0	6.4
Arylsulfatase	159	Arylsulfatase	29.4	21.3	23.8	6.8	24.3	11.1	33.6	5.6
Arylsulfatase	160	Arylsulfatase	18.6	9.9	14.3	10.8	17.8	7.0	20.6	6.1
Arylsulfatase	161	Arylsulfatase	25.5	9.9	22.8	18.3	23.7	16.9	27.6	13.5
Arylsulfatase	162	Arylsulfatase	8.8	5.6	9.6	7.7	11.2	5.5	11.2	3.4
Arylsulfatase	163	Arylsulfatase	60.6	40.2	45.5	37.9	31.3	19.3	42.8	24.1
Arylsulfatase	164	Arylsulfatase	12.8	7.1	12.9	14.4	13.9	11.5	17.7	6.4
Arylsulfatase	165	Arylsulfatase	11.3	10.3	11.1	12.4	18.8	17.5	15.7	14.1
Arylsulfatase	166	Arylsulfatase	35.9	28.2	39.5	32.2	40.8	24.7	43.3	19.2
Arylsulfatase	167	Arylsulfatase	51.5	23.3	30.3	16.6	22.4	15.9	27.6	13.0
Arylsulfatase	168	Arylsulfatase	124.1	47.4	84.7	57.9	84.6	39.1	96.8	32.8
Arylsulfatase	169	Arylsulfatase	18.3	17.3	9.9	7.1	19.2	9.9	18.4	12.6
Arylsulfatase	170	Arylsulfatase	19.3	20.1	15.0	17.4	24.0	19.8	28.6	23.0
Arylsulfatase	171	Arylsulfatase	12.9	11.3	6.2	6.6	11.8	9.0	13.9	9.9
Arylsulfatase	172	Arylsulfatase	10.0	5.1	8.7	7.5	10.1	10.0	14.7	9.0
Arylsulfatase	173	Arylsulfatase	3.0	4.1	1.4	0.9	2.0	2.8	3.2	3.7
Arylsulfatase	174	Arylsulfatase	52.6	19.5	40.5	27.9	44.1	28.0	45.3	19.6
Arylsulfatase	175	Arylsulfatase	8.5	3.5	3.9	2.6	4.5	4.4	3.6	3.2
Arylsulfatase	176	Arylsulfatase	27.0	15.4	15.3	9.2	17.4	14.7	12.9	6.5
Arylsulfatase	177	Arylsulfatase	18.5	7.0	18.5	15.1	18.1	11.7	16.7	5.5

Arylsulfatase	178	Arylsulfatase	28.1	14.5	23.4	13.0	23.3	9.3	24.2	11.3
Arylsulfatase	179	Arylsulfatase	110.4	72.2	105.2	44.7	55.8	31.6	70.6	27.5
Arylsulfatase	180	Arylsulfatase	15.8	12.6	12.0	11.5	19.3	9.5	20.8	10.7
Arylsulfatase	181	Arylsulfatase	70.0	34.4	57.6	43.5	59.0	33.0	58.8	33.2
Arylsulfatase	182	Arylsulfatase	7.6	5.0	9.2	6.9	17.8	13.4	15.2	7.2
Arylsulfatase	183	Arylsulfatase	22.2	10.8	20.6	16.1	39.5	29.5	36.1	20.0
Arylsulfatase	184	Arylsulfatase	39.4	24.5	34.6	25.4	40.6	18.1	52.6	27.4
Cell division	1	Cell division inhibitor	183.6	27.7	119.8	40.2	119.4	44.7	130.6	37.6
Cell division	2	Cell division inhibitor, YfcH	39.1	20.6	50.2	18.7	73.0	26.5	77.0	29.1
Cell division	3	FtsA	85.9	78.4	38.7	13.0	73.9	30.5	63.9	14.5
Cell division	4	FtsH	636.6	204.2	507.0	207.6	427.8	204.1	459.4	247.6
Cell division	5	FtsH	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cell division	6	FtsI	313.1	178.9	245.4	203.9	249.0	176.4	267.9	156.5
Cell division	7	FtsW	106.2	68.1	65.4	42.5	83.5	36.0	95.9	30.9
Cell division	8	FtsZ	32.9	12.9	23.3	10.8	38.2	12.4	39.6	15.6
Cell division	9	Chromosomal replication initiator protein, DnaA	235.3	118.1	176.0	37.4	145.3	44.8	173.4	63.5
Cell division	10	Chromosome partitioning protein, ParA	133.8	81.7	85.6	40.6	75.8	36.9	80.6	27.7
Cell division	11	Chromosome partitioning protein, ParB	82.2	21.8	42.4	17.1	38.6	10.1	59.1	26.9
Cell division	12	Chromosome partition protein, smc	47.6	27.0	27.1	14.8	24.2	10.3	28.3	14.7
Cell division	13	Rod shape-determining protein, MreB	335.7	395.2	156.4	33.3	157.2	132.9	162.5	128.7
Cell division	14	Rod shape-determining protein, RodA	249.6	247.8	96.2	35.2	70.7	50.7	76.9	65.9
Cell division	15	Circadian phase modifier	128.8	73.1	122.9	99.2	122.2	77.0	140.3	77.9
Fucose	1	alpha-L-fucosidase	46.7	23.4	48.7	35.2	63.1	31.3	65.6	47.1
Fucose	2	alpha-L-fucosidase	134.5	87.3	148.1	133.3	188.5	114.3	200.8	114.9
Fucose	3	alpha-L-fucosidase	52.4	51.9	29.9	25.6	26.8	18.4	33.7	20.6

Fucose	4	alpha-L-fucosidase	94.9	51.7	70.9	46.0	67.7	34.8	70.2	27.2
Fucose	5	alpha-L-fucosidase	86.8	32.7	126.9	53.2	123.4	66.6	122.5	56.6
Fucose	6	alpha-L-fucosidase	25.4	12.3	18.9	9.9	24.2	11.0	20.0	6.2
Fucose	7	alpha-L-fucosidase	31.5	18.9	25.9	19.9	14.7	8.0	11.7	12.3
Fucose	8	alpha-L-fucosidase	113.6	62.9	105.6	79.1	121.8	53.2	114.2	46.9
Fucose	9	alpha-L-fucosidase	11.6	9.7	11.6	7.2	12.9	5.5	19.4	11.7
Fucose	10	alpha-L-fucosidase	21.0	17.1	21.4	19.7	24.5	16.4	18.0	11.0
Fucose	11	alpha-L-fucosidase	70.1	30.5	58.2	27.9	60.1	18.7	65.0	16.9
Fucose	12	alpha-L-fucosidase	39.5	16.7	32.4	24.2	36.0	33.5	37.2	20.0
Fucose	13	alpha-L-fucosidase	55.2	29.0	45.8	12.9	60.4	22.1	71.2	16.0
Fucose	14	alpha-L-fucosidase	20.3	14.2	13.7	12.4	17.7	14.9	21.3	22.5
Fucose	15	alpha-L-fucosidase	331.3	266.0	286.5	214.8	569.1	388.1	736.2	603.0
Fucose	16	alpha-L-fucosidase	37.5	23.1	21.6	10.4	32.6	19.1	39.6	16.4
Fucose	17	alpha-L-fucosidase	15.5	11.5	12.5	9.4	20.3	9.3	20.8	13.4
Fucose	18	alpha-L-fucosidase	11.3	8.4	11.6	10.6	8.5	5.4	13.4	8.9
Fucose	19	alpha-L-fucosidase	19.8	14.7	8.4	7.6	8.9	5.3	9.0	5.8
Fucose	20	alpha-L-fucosidase	11.1	7.7	11.0	10.5	15.4	7.0	19.1	11.9
Fucose	21	alpha-L-fucosidase	55.3	50.5	28.2	30.7	25.2	12.8	35.3	13.7
Fucose	22	alpha-L-fucosidase	117.8	130.9	61.7	83.3	38.5	20.3	43.7	30.8
Fucose	23	alpha-L-fucosidase	28.6	13.6	27.0	18.1	28.2	19.7	32.9	20.2
Fucose	24	alpha-L-fucosidase	78.2	36.2	36.1	23.3	21.3	12.3	28.9	13.5
Fucose	25	Fucose permease	133.7	62.7	103.1	48.9	93.5	37.3	105.2	41.3
Fucose	26	Fucose permease	23.7	10.8	28.1	7.4	26.1	9.2	26.1	5.6
Fucose	27	L-fucose isomerase	408.4	211.8	228.6	232.4	92.0	56.2	121.6	64.2
Fucose	28	L-fucose kinase	38.0	12.5	36.4	8.8	27.6	10.3	33.4	11.7
Fucose	29	L-fucose mutarotase	42.0	10.7	37.7	16.2	44.9	19.4	44.6	17.1

Fucose	30	GDP-L-fucose synthetase	26.3	14.9	13.2	7.0	21.1	8.0	24.5	13.7
Glycoside hydrolases	1	GH109	102.1	54.6	65.6	39.7	85.3	47.7	87.8	52.0
Glycoside hydrolases	2	GH117	22.3	12.2	21.2	12.0	24.3	15.3	23.9	12.1
Glycoside hydrolases	3	GH78	30.1	21.4	22.0	15.3	18.0	15.0	21.8	18.4
Glycoside hydrolases	4	GH82	25.7	10.5	13.2	6.9	21.7	7.2	18.2	7.3
Glycoside hydrolases	5	GH0	21.9	11.6	18.4	6.6	22.2	8.6	20.3	9.0
Glycoside hydrolases	6	GH0	52.7	18.0	46.2	15.9	104.4	101.7	80.0	40.7
Glycoside hydrolases	7	GH0	18.2	9.0	25.3	11.4	14.1	7.0	18.4	8.2
Glycoside hydrolases	8	GH10	18.0	7.9	18.5	11.6	21.3	9.8	27.7	16.3
Glycoside hydrolases	9	GH10	39.2	18.9	31.4	15.5	45.8	22.3	41.8	11.5
Glycoside hydrolases	10	GH10	18.3	10.5	12.2	11.6	18.9	13.0	26.1	15.8
Glycoside hydrolases	11	GH10	30.3	14.5	24.8	7.1	29.0	27.9	26.2	11.1
Glycoside hydrolases	12	GH10	40.3	37.0	19.9	22.3	13.0	8.6	18.0	26.8
Glycoside hydrolases	13	GH10	90.1	116.6	12.8	13.8	28.6	38.2	26.6	28.6
Glycoside hydrolases	14	GH10	57.0	65.1	19.9	32.7	23.5	31.9	34.1	58.0
Glycoside hydrolases	15	GH10	38.9	25.3	65.2	38.1	50.4	43.8	55.4	33.4
Glycoside hydrolases	16	GH10	97.0	37.5	101.5	43.0	63.2	31.1	77.6	29.9
Glycoside hydrolases	17	GH10	31.6	15.9	29.9	15.3	29.5	17.6	31.6	14.6
Glycoside hydrolases	18	GH101	93.5	101.5	27.0	33.2	22.4	27.8	24.9	30.8
Glycoside hydrolases	19	GH106	25.4	34.9	8.9	10.3	5.6	7.0	10.3	16.1

Glycoside hydrolases	20	GH106	42.3	18.6	51.6	39.8	48.7	29.3	75.2	35.2
Glycoside hydrolases	21	GH107	41.5	44.6	13.2	16.3	14.6	19.6	20.0	29.6
Glycoside hydrolases	22	GH109	34.9	43.3	9.8	9.5	10.8	16.9	12.2	9.8
Glycoside hydrolases	23	GH109	128.1	39.0	77.9	45.6	130.8	75.9	144.1	109.4
Glycoside hydrolases	24	GH109	22.8	8.5	16.6	6.3	22.7	7.6	31.2	11.2
Glycoside hydrolases	25	GH109	14.5	10.4	8.0	4.4	14.1	8.2	10.5	5.2
Glycoside hydrolases	26	GH109	485.8	255.9	787.5	529.8	626.8	401.3	543.1	325.8
Glycoside hydrolases	27	GH109	25.0	7.9	23.7	12.0	24.2	10.5	22.6	8.3
Glycoside hydrolases	28	GH109	32.5	8.5	33.3	7.1	30.5	17.5	22.7	9.5
Glycoside hydrolases	29	GH109	33.4	31.1	13.0	1.7	24.2	15.8	16.5	17.4
Glycoside hydrolases	30	GH109	132.3	67.4	124.9	81.9	144.7	94.8	154.9	97.2
Glycoside hydrolases	31	GH109	82.0	43.4	80.7	39.4	62.7	26.5	72.0	29.3
Glycoside hydrolases	32	GH109	7.5	5.7	8.0	4.8	8.3	5.5	7.9	4.8
Glycoside hydrolases	33	GH109	8.6	6.0	7.2	5.9	13.0	6.6	11.6	4.3
Glycoside hydrolases	34	GH109	40.7	17.2	40.0	22.3	41.3	19.9	48.5	21.4
Glycoside hydrolases	35	GH109	30.7	17.7	29.3	9.5	37.0	11.7	39.0	7.5
Glycoside hydrolases	36	GH109	48.1	28.0	55.0	20.6	46.7	30.6	60.1	20.1
Glycoside hydrolases	37	GH109	81.8	34.3	101.5	42.6	68.5	53.2	76.8	39.2
Glycoside hydrolases	38	GH109	9.0	4.0	8.0	6.3	13.0	11.0	11.2	6.5

Glycoside hydrolases	39	GH109	59.8	26.2	60.4	34.7	93.2	57.1	82.4	36.3
Glycoside hydrolases	40	GH109	21.3	6.9	13.1	4.0	17.1	6.6	16.3	5.0
Glycoside hydrolases	41	GH109	12.8	15.5	7.4	4.4	6.3	5.1	9.8	7.4
Glycoside hydrolases	42	GH109	116.8	31.1	122.7	53.4	59.8	25.2	65.9	52.7
Glycoside hydrolases	43	GH109	9.3	8.1	6.6	5.4	8.0	5.0	8.7	9.2
Glycoside hydrolases	44	GH109	27.4	13.1	19.8	17.2	29.0	17.1	34.4	15.2
Glycoside hydrolases	45	GH109	91.7	80.5	149.8	145.6	134.6	131.1	118.5	107.2
Glycoside hydrolases	46	GH109	90.6	24.3	66.2	18.2	64.4	19.5	72.0	25.6
Glycoside hydrolases	47	GH109	98.0	54.8	67.0	37.7	71.1	55.4	78.7	46.4
Glycoside hydrolases	48	GH109	476.4	236.2	335.4	101.6	297.0	52.6	301.7	103.5
Glycoside hydrolases	49	GH109	25.2	17.1	18.6	11.6	18.5	9.9	20.0	7.5
Glycoside hydrolases	50	GH109	56.8	48.6	27.8	25.0	45.2	24.3	40.4	23.8
Glycoside hydrolases	51	GH109	101.5	89.2	43.5	38.7	53.0	34.2	45.1	24.2
Glycoside hydrolases	52	GH109	48.2	52.8	29.1	23.5	35.7	23.5	31.2	16.4
Glycoside hydrolases	53	GH109	63.4	43.3	51.0	48.0	46.9	41.2	60.1	40.8
Glycoside hydrolases	54	GH109	45.5	29.1	51.0	31.7	24.3	11.1	25.9	9.0
Glycoside hydrolases	55	GH109	16.8	5.6	12.6	10.0	13.6	7.7	12.1	4.9
Glycoside hydrolases	56	GH109	20.6	17.4	16.0	9.1	15.3	11.2	14.3	10.6
Glycoside hydrolases	57	GH109	17.8	15.3	11.1	9.1	18.0	12.2	24.9	9.0

Glycoside hydrolases	58	GH110	20.6	11.4	25.7	9.6	26.1	10.0	28.2	14.8
Glycoside hydrolases	59	GH110	57.6	21.3	63.1	33.2	57.6	19.4	65.5	22.6
Glycoside hydrolases	60	GH110	14.8	11.4	10.1	4.6	20.6	12.5	11.1	8.0
Glycoside hydrolases	61	GH110	58.2	29.0	44.0	19.3	51.5	20.2	50.2	25.3
Glycoside hydrolases	62	GH115	367.5	414.2	137.7	82.5	150.4	73.5	114.7	53.4
Glycoside hydrolases	63	GH116	495.4	316.7	141.0	43.5	178.0	102.0	178.4	91.9
Glycoside hydrolases	64	GH116	62.5	51.2	163.4	218.1	49.4	20.6	63.8	48.7
Glycoside hydrolases	65	GH117	37.8	30.7	30.3	16.5	24.6	9.0	16.9	7.6
Glycoside hydrolases	66	GH117	174.5	85.9	167.5	75.4	193.8	88.3	177.7	77.6
Glycoside hydrolases	67	GH117 (Top 10: row 7)	756.0	897.9	290.2	200.3	454.3	333.9	375.9	249.1
Glycoside hydrolases	68	GH117	155.4	57.4	157.9	38.1	229.3	125.0	196.9	105.1
Glycoside hydrolases	69	GH117	136.7	81.9	60.2	17.9	108.3	62.9	118.0	69.4
Glycoside hydrolases	70	GH117	129.0	24.0	76.9	21.0	145.0	126.8	119.3	61.6
Glycoside hydrolases	71	GH117	70.0	61.8	27.4	6.8	52.8	37.4	42.9	31.9
Glycoside hydrolases	72	GH117	629.1	228.1	504.2	275.2	288.0	123.9	336.9	168.4
Glycoside hydrolases	73	GH117	61.9	21.3	50.0	18.8	43.0	26.4	45.9	16.5
Glycoside hydrolases	74	GH117	89.9	28.6	58.3	39.3	79.5	34.7	80.3	24.7
Glycoside hydrolases	75	GH117	451.1	230.0	460.1	343.9	311.1	285.7	342.6	288.3
Glycoside hydrolases	76	GH117 (Top 10: row 6)	867.6	521.8	809.3	458.2	439.8	330.6	454.6	331.3

Glycoside hydrolases	77	GH117 (Top 10: row 1)	17327.2	9212.3	21467.5	10974.4	8483.0	6399.4	9061.0	6754.6
Glycoside hydrolases	78	GH117	496.3	331.8	504.5	134.6	351.1	164.2	357.6	173.4
Glycoside hydrolases	79	GH117	152.7	64.1	122.4	63.6	137.4	57.2	118.2	43.0
Glycoside hydrolases	80	GH117	331.2	118.4	313.2	159.8	382.4	69.5	382.5	55.0
Glycoside hydrolases	81	GH117	17.1	11.7	14.2	9.1	17.5	13.5	22.0	9.7
Glycoside hydrolases	82	GH120	53.3	30.3	34.5	10.5	27.6	12.5	32.1	15.6
Glycoside hydrolases	83	GH127	32.1	18.7	18.9	12.1	28.9	18.9	38.5	15.9
Glycoside hydrolases	84	GH127	27.9	15.8	19.6	13.8	25.6	14.1	25.9	16.0
Glycoside hydrolases	85	GH127	11.5	5.9	10.7	6.8	18.6	8.3	19.5	4.9
Glycoside hydrolases	86	GH127	6.8	3.5	6.0	5.1	9.3	7.7	10.4	4.2
Glycoside hydrolases	87	GH127	24.1	12.5	15.3	12.4	29.9	20.7	29.3	20.6
Glycoside hydrolases	88	GH127	39.1	16.6	24.0	8.0	19.0	9.3	17.5	5.1
Glycoside hydrolases	89	GH128	72.5	66.5	51.6	38.6	53.8	31.0	52.5	38.8
Glycoside hydrolases	90	GH129	22.8	13.7	30.2	25.5	21.7	22.1	23.8	19.4
Glycoside hydrolases	91	GH13	40.1	12.1	22.9	10.3	31.4	12.8	31.5	10.8
Glycoside hydrolases	92	GH13	138.6	65.4	114.3	66.0	120.8	128.1	113.7	107.4
Glycoside hydrolases	93	GH13	121.9	124.4	50.2	24.2	60.8	42.5	62.9	35.2
Glycoside hydrolases	94	GH13	39.3	24.4	38.1	33.7	28.4	25.8	28.6	28.5
Glycoside hydrolases	95	GH13	37.2	14.0	27.5	8.2	26.4	8.3	24.9	10.1

Glycoside hydrolases	96	GH13	47.3	28.7	60.7	36.2	87.9	45.4	98.5	45.5
Glycoside hydrolases	97	GH13	225.9	253.3	102.4	92.1	120.8	78.1	118.2	44.9
Glycoside hydrolases	98	GH13 (Top 10: row 9)	379.5	271.3	434.8	344.3	601.9	266.7	627.1	276.8
Glycoside hydrolases	99	GH130	181.4	56.9	199.6	137.1	162.8	90.4	200.6	97.0
Glycoside hydrolases	100	GH130	28.3	24.5	28.4	17.5	24.0	13.3	31.3	11.9
Glycoside hydrolases	101	GH130	19.4	12.2	17.9	13.7	24.5	10.2	25.4	14.4
Glycoside hydrolases	102	GH130	17.4	11.2	18.6	15.4	28.2	25.6	22.6	7.9
Glycoside hydrolases	103	GH136	209.5	133.6	187.9	83.1	132.0	55.2	159.5	71.5
Glycoside hydrolases	104	GH136	25.9	12.4	13.7	9.8	11.7	11.0	15.6	12.6
Glycoside hydrolases	105	GH136	18.9	15.1	12.7	10.0	13.9	2.8	11.4	5.5
Glycoside hydrolases	106	GH138	32.3	17.9	45.8	16.4	36.9	16.3	49.9	13.9
Glycoside hydrolases	107	GH138	64.1	42.1	82.4	43.2	59.3	31.1	53.1	18.4
Glycoside hydrolases	108	GH138	60.5	27.5	53.0	32.4	47.5	26.5	50.0	28.2
Glycoside hydrolases	109	GH138	80.1	49.8	77.3	67.2	51.6	33.1	42.7	41.5
Glycoside hydrolases	110	GH138	87.0	67.2	56.1	14.5	46.3	15.5	56.0	20.8
Glycoside hydrolases	111	GH14	109.5	88.3	85.5	69.7	70.2	46.4	81.8	59.7
Glycoside hydrolases	112	GH140	32.2	11.8	25.8	12.5	32.3	12.4	29.1	12.7
Glycoside hydrolases	113	GH140	44.0	14.6	29.4	12.7	36.0	12.7	47.6	14.8
Glycoside hydrolases	114	GH140	21.5	8.5	16.3	7.6	17.7	10.0	15.1	12.1

Glycoside hydrolases	115	GH141	6.8	4.9	6.0	4.1	5.5	5.6	7.3	5.6
Glycoside hydrolases	116	GH141	26.2	9.5	12.8	10.1	19.9	18.3	9.0	7.5
Glycoside hydrolases	117	GH141	16.6	7.7	16.0	12.5	21.8	10.4	22.1	18.4
Glycoside hydrolases	118	GH141	19.9	7.7	19.5	8.7	20.6	11.1	27.7	15.6
Glycoside hydrolases	119	GH141	20.4	13.4	16.3	10.6	17.3	15.9	23.8	15.6
Glycoside hydrolases	120	GH141	33.3	15.7	25.0	19.4	30.0	12.7	34.1	17.2
Glycoside hydrolases	121	GH141	20.0	9.4	13.0	6.8	36.3	23.3	41.7	24.5
Glycoside hydrolases	122	GH141	20.5	12.1	28.0	10.7	16.0	7.7	17.4	9.0
Glycoside hydrolases	123	GH148	176.9	155.3	56.3	41.1	60.8	58.6	51.8	32.8
Glycoside hydrolases	124	GH148	83.6	80.5	23.6	17.5	19.1	24.0	26.2	32.6
Glycoside hydrolases	125	GH149	31.0	24.8	25.7	25.3	12.0	9.8	19.8	18.3
Glycoside hydrolases	126	GH150	0.3	0.7	2.6	5.2	0.0	0.0	0.0	0.0
Glycoside hydrolases	127	GH150	33.2	20.9	19.6	12.3	10.6	6.0	18.3	14.8
Glycoside hydrolases	128	GH151	31.3	23.9	16.5	8.6	16.8	8.5	11.1	6.6
Glycoside hydrolases	129	GH151	15.3	10.6	10.0	7.3	9.4	4.8	12.4	8.1
Glycoside hydrolases	130	GH151	18.8	16.2	12.3	6.0	13.0	8.7	14.5	6.0
Glycoside hydrolases	131	GH151	17.5	13.9	10.5	5.7	9.4	4.9	10.8	5.2
Glycoside hydrolases	132	GH16	37.0	15.1	32.0	11.3	30.4	14.7	25.9	18.3
Glycoside hydrolases	133	GH16	38.5	23.9	27.3	18.4	47.4	25.7	48.4	29.1

Glycoside hydrolases	134	GH16	45.8	17.3	28.5	15.5	47.4	18.8	52.7	20.5
Glycoside hydrolases	135	GH16	59.2	25.8	34.0	21.2	63.7	23.4	61.7	23.2
Glycoside hydrolases	136	GH16	202.7	106.1	366.8	106.3	390.8	77.4	482.2	72.9
Glycoside hydrolases	137	GH16	85.5	39.6	61.6	43.7	72.4	28.3	63.1	28.0
Glycoside hydrolases	138	GH16 (Top 10: row 3)	1260.1	468.8	1135.7	457.3	576.2	289.8	620.3	348.6
Glycoside hydrolases	139	GH16	12.4	7.7	10.1	7.5	11.4	8.4	12.0	3.9
Glycoside hydrolases	140	GH16	24.4	11.6	18.1	13.2	24.3	16.2	29.5	11.5
Glycoside hydrolases	141	GH16	23.3	22.9	18.9	11.0	47.5	35.7	47.3	37.3
Glycoside hydrolases	142	GH16	20.3	15.2	16.1	12.0	10.3	7.3	15.2	11.4
Glycoside hydrolases	143	GH16	47.3	35.3	42.1	28.1	33.0	17.7	31.8	17.9
Glycoside hydrolases	144	GH16	20.7	19.0	13.2	8.2	11.1	7.4	13.1	10.1
Glycoside hydrolases	145	GH16	45.8	31.9	44.9	31.0	27.1	16.6	32.3	21.8
Glycoside hydrolases	146	GH17	32.4	21.1	17.9	12.5	17.0	10.1	16.6	10.8
Glycoside hydrolases	147	GH17	23.4	13.9	24.2	12.1	28.4	12.6	25.2	10.1
Glycoside hydrolases	148	GH2	52.6	43.1	34.1	36.5	29.3	29.3	30.0	27.0
Glycoside hydrolases	149	GH2	31.3	13.6	28.8	13.1	26.6	11.5	32.9	12.7
Glycoside hydrolases	150	GH2	26.1	11.7	23.6	22.0	28.4	16.7	27.4	13.8
Glycoside hydrolases	151	GH2	30.9	17.0	32.4	18.0	40.4	16.3	40.9	11.8
Glycoside hydrolases	152	GH2	48.5	34.8	35.6	16.1	23.4	9.1	28.3	14.7

Glycoside hydrolases	153	GH2	114.6	91.2	102.5	69.0	58.2	35.0	52.0	31.1
Glycoside hydrolases	154	GH2	23.8	13.3	13.5	11.3	13.9	7.5	11.4	6.4
Glycoside hydrolases	155	GH2	13.6	12.4	11.1	7.5	13.4	11.0	15.1	7.7
Glycoside hydrolases	156	GH2	35.9	25.2	38.6	29.3	38.8	25.7	35.5	21.9
Glycoside hydrolases	157	GH2	7.1	4.7	7.4	6.5	9.4	5.3	8.5	3.4
Glycoside hydrolases	158	GH2	12.5	8.2	10.0	8.9	20.8	17.6	25.9	25.1
Glycoside hydrolases	159	GH2	30.8	22.2	18.8	14.7	62.1	59.6	62.9	53.0
Glycoside hydrolases	160	GH2	133.2	120.6	85.0	94.9	228.8	166.1	292.4	278.3
Glycoside hydrolases	161	GH2	33.1	12.8	38.7	27.4	39.6	23.6	33.0	15.7
Glycoside hydrolases	162	GH2	50.5	20.1	45.7	26.5	34.1	15.5	37.9	18.2
Glycoside hydrolases	163	GH2	13.0	10.7	8.5	3.3	16.7	16.2	13.9	13.9
Glycoside hydrolases	164	GH2	24.9	21.4	21.4	18.6	24.2	11.1	24.9	10.1
Glycoside hydrolases	165	GH2	18.9	7.4	18.6	11.6	21.9	9.7	21.9	8.9
Glycoside hydrolases	166	GH2	5.7	3.9	3.0	2.1	4.6	2.6	5.6	3.1
Glycoside hydrolases	167	GH2	10.9	10.6	14.1	10.8	16.2	7.1	16.2	11.9
Glycoside hydrolases	168	GH2	28.3	14.1	20.8	15.9	25.4	12.4	36.0	16.5
Glycoside hydrolases	169	GH2	39.7	24.3	27.7	18.9	14.8	9.8	18.2	10.3
Glycoside hydrolases	170	GH2	8.0	6.8	6.9	6.5	4.4	3.8	4.4	5.7
Glycoside hydrolases	171	GH20	62.2	43.1	80.3	65.1	88.3	50.0	98.7	50.8

Glycoside hydrolases	172	GH26	5.6	4.5	6.1	2.2	3.0	3.7	7.5	10.1
Glycoside hydrolases	173	GH26	10.0	5.8	9.0	6.1	9.4	7.6	8.8	6.9
Glycoside hydrolases	174	GH27	20.0	6.8	19.2	13.5	22.6	15.7	26.8	13.6
Glycoside hydrolases	175	GH27	8.6	9.7	6.3	4.7	7.9	8.0	10.9	9.8
Glycoside hydrolases	176	GH28	42.5	29.7	28.3	14.6	20.9	19.7	14.7	9.0
Glycoside hydrolases	177	GH28	52.9	26.4	43.3	21.5	51.3	21.0	50.6	20.6
Glycoside hydrolases	178	GH28	33.0	23.7	23.3	15.6	21.3	20.5	22.7	15.4
Glycoside hydrolases	179	GH28	23.7	22.3	12.9	11.2	11.5	7.6	17.1	13.5
Glycoside hydrolases	180	GH28	21.7	15.0	18.8	15.0	29.2	21.5	28.6	24.2
Glycoside hydrolases	181	GH28	112.1	95.1	48.8	18.5	57.3	20.8	64.0	19.6
Glycoside hydrolases	182	GH29	46.7	23.4	48.7	35.2	63.1	31.3	65.6	47.1
Glycoside hydrolases	183	GH29	134.5	87.3	148.1	133.3	188.5	114.3	200.8	114.9
Glycoside hydrolases	184	GH29	52.4	51.9	29.9	25.6	26.8	18.4	33.7	20.6
Glycoside hydrolases	185	GH29	94.9	51.7	70.9	46.0	67.7	34.8	70.2	27.2
Glycoside hydrolases	186	GH29	86.8	32.7	126.9	53.2	123.4	66.6	122.5	56.6
Glycoside hydrolases	187	GH29	25.4	12.3	18.9	9.9	24.2	11.0	20.0	6.2
Glycoside hydrolases	188	GH29	31.5	18.9	25.9	19.9	14.7	8.0	11.7	12.3
Glycoside hydrolases	189	GH29	113.6	62.9	105.6	79.1	121.8	53.2	114.2	46.9
Glycoside hydrolases	190	GH29	11.6	9.7	11.6	7.2	12.9	5.5	19.4	11.7

Glycoside hydrolases	191	GH29	21.0	17.1	21.4	19.7	24.5	16.4	18.0	11.0
Glycoside hydrolases	192	GH29	70.1	30.5	58.2	27.9	60.1	18.7	65.0	16.9
Glycoside hydrolases	193	GH29	39.5	16.7	32.4	24.2	36.0	33.5	37.2	20.0
Glycoside hydrolases	194	GH29	55.2	29.0	45.8	12.9	60.4	22.1	71.2	16.0
Glycoside hydrolases	195	GH29	20.3	14.2	13.7	12.4	17.7	14.9	21.3	22.5
Glycoside hydrolases	196	GH29	331.3	266.0	286.5	214.8	569.1	388.1	736.2	603.0
Glycoside hydrolases	197	GH29	37.5	23.1	21.6	10.4	32.6	19.1	39.6	16.4
Glycoside hydrolases	198	GH29	15.5	11.5	12.5	9.4	20.3	9.3	20.8	13.4
Glycoside hydrolases	199	GH29	11.3	8.4	11.6	10.6	8.5	5.4	13.4	8.9
Glycoside hydrolases	200	GH29	19.8	14.7	8.4	7.6	8.9	5.3	9.0	5.8
Glycoside hydrolases	201	GH29	11.1	7.7	11.0	10.5	15.4	7.0	19.1	11.9
Glycoside hydrolases	202	GH29	55.3	50.5	28.2	30.7	25.2	12.8	35.3	13.7
Glycoside hydrolases	203	GH29	117.8	130.9	61.7	83.3	38.5	20.3	43.7	30.8
Glycoside hydrolases	204	GH29	28.6	13.6	27.0	18.1	28.2	19.7	32.9	20.2
Glycoside hydrolases	205	GH29	78.2	36.2	36.1	23.3	21.3	12.3	28.9	13.5
Glycoside hydrolases	206	GH29	13.9	11.6	12.2	12.2	13.0	12.9	18.6	17.1
Glycoside hydrolases	207	GH29	7.2	6.1	5.9	5.5	1.1	2.4	0.8	1.8
Glycoside hydrolases	208	GH3	12.9	9.6	9.1	3.7	12.2	6.9	17.0	5.4
Glycoside hydrolases	209	GH3	64.9	24.3	42.5	18.4	57.5	25.5	48.5	14.6

Glycoside hydrolases	210	GH30	37.3	23.9	25.4	19.3	19.9	17.3	18.8	13.5
Glycoside hydrolases	211	GH30	34.8	20.8	22.8	17.2	17.0	13.9	16.3	10.5
Glycoside hydrolases	212	GH30	9.6	6.2	8.5	5.9	13.1	7.0	16.1	10.5
Glycoside hydrolases	213	GH31	24.7	14.3	30.8	7.9	29.1	14.7	34.7	6.8
Glycoside hydrolases	214	GH31	10.7	13.3	11.1	6.3	13.7	11.4	14.2	11.7
Glycoside hydrolases	215	GH31	27.1	14.9	24.7	17.5	19.1	9.8	20.5	12.0
Glycoside hydrolases	216	GH31	11.7	13.6	6.8	5.6	3.9	4.6	9.4	4.1
Glycoside hydrolases	217	GH32	16.9	16.1	10.1	9.1	13.7	10.5	13.1	9.8
Glycoside hydrolases	218	GH32	24.4	19.0	17.7	17.6	22.2	20.6	26.3	24.6
Glycoside hydrolases	219	GH33	40.7	29.6	27.7	25.7	42.5	48.3	36.6	38.1
Glycoside hydrolases	220	GH33	30.7	17.2	27.4	26.1	36.2	17.9	41.3	12.4
Glycoside hydrolases	221	GH33	42.9	30.5	24.9	16.0	30.8	14.1	30.0	15.9
Glycoside hydrolases	222	GH33	25.5	14.9	24.3	10.6	28.0	10.6	22.6	14.7
Glycoside hydrolases	223	GH33	14.1	7.0	10.2	5.8	13.3	8.3	13.7	5.2
Glycoside hydrolases	224	GH33	7.8	6.1	4.3	5.7	6.1	7.9	16.3	17.8
Glycoside hydrolases	225	GH33	6.8	6.1	5.8	3.4	11.0	4.8	13.0	7.3
Glycoside hydrolases	226	GH33	43.7	34.2	34.0	20.3	21.8	12.8	21.8	12.2
Glycoside hydrolases	227	GH33	32.2	21.1	24.3	13.0	22.0	12.3	20.4	10.1
Glycoside hydrolases	228	GH33	27.5	19.9	18.1	12.7	18.6	15.5	16.4	11.9

Glycoside hydrolases	229	GH33	37.0	13.5	36.6	12.3	22.6	10.2	23.8	6.1
Glycoside hydrolases	230	GH33	25.2	16.5	19.6	13.4	28.8	15.4	24.4	15.9
Glycoside hydrolases	231	GH36	39.0	17.0	35.5	18.2	27.1	19.9	33.2	21.4
Glycoside hydrolases	232	GH36	19.2	12.4	7.4	3.0	11.9	8.7	13.8	3.4
Glycoside hydrolases	233	GH36	8.2	10.3	5.9	5.8	5.5	7.6	3.1	2.8
Glycoside hydrolases	234	GH38	17.3	9.8	22.2	11.7	31.4	20.2	24.7	13.5
Glycoside hydrolases	235	GH38	14.6	7.5	20.8	11.1	21.9	9.0	24.5	11.0
Glycoside hydrolases	236	GH39	28.9	24.0	19.4	12.4	28.9	26.8	30.1	34.1
Glycoside hydrolases	237	GH39	20.1	22.5	12.8	14.8	25.1	22.3	16.2	12.2
Glycoside hydrolases	238	GH39	21.1	14.7	5.2	4.9	15.7	14.1	12.2	6.4
Glycoside hydrolases	239	GH4	8.6	5.2	11.1	8.7	11.3	7.9	11.7	12.9
Glycoside hydrolases	240	GH4	36.7	18.5	41.6	26.0	95.0	48.2	116.3	59.8
Glycoside hydrolases	241	GH42	28.2	17.2	34.7	24.6	53.8	39.8	52.0	35.7
Glycoside hydrolases	242	GH42	30.8	19.9	18.0	13.0	11.6	7.6	13.0	8.6
Glycoside hydrolases	243	GH43	80.9	75.0	91.4	63.4	55.4	47.2	57.5	50.7
Glycoside hydrolases	244	GH43	31.6	14.6	25.1	15.1	23.8	15.5	30.6	14.7
Glycoside hydrolases	245	GH43	16.3	13.9	14.1	13.0	19.4	12.2	23.1	16.3
Glycoside hydrolases	246	GH43	25.0	17.2	20.8	15.0	24.2	10.6	24.7	10.6
Glycoside hydrolases	247	GH43	13.1	9.6	14.0	9.6	13.1	6.3	11.5	6.7

Glycoside hydrolases	248	GH43	14.3	13.0	11.3	8.1	18.2	17.7	20.6	12.5
Glycoside hydrolases	249	GH43	32.9	20.4	12.5	8.2	45.1	64.3	58.1	98.3
Glycoside hydrolases	250	GH43	68.5	46.0	33.1	22.3	77.9	69.7	53.2	52.7
Glycoside hydrolases	251	GH43	18.8	7.8	15.5	12.1	17.8	10.1	17.7	8.5
Glycoside hydrolases	252	GH49	61.1	48.3	37.3	30.4	45.5	22.5	35.4	23.0
Glycoside hydrolases	253	GH5	21.5	14.1	22.8	12.8	16.7	11.2	27.5	17.8
Glycoside hydrolases	254	GH5	153.4	123.4	113.5	106.4	61.5	76.5	50.7	41.2
Glycoside hydrolases	255	GH5	30.3	22.2	20.1	20.9	27.7	24.0	32.9	30.4
Glycoside hydrolases	256	GH5	31.4	25.5	20.7	8.0	20.7	13.1	26.6	18.7
Glycoside hydrolases	257	GH5	64.0	51.3	42.4	38.8	116.1	70.4	123.8	105.3
Glycoside hydrolases	258	GH50	43.2	28.5	30.1	15.9	82.5	44.6	81.8	43.7
Glycoside hydrolases	259	GH50	29.4	19.8	11.8	10.5	20.0	18.0	17.4	15.3
Glycoside hydrolases	260	GH50	76.3	37.0	79.9	50.0	110.9	54.3	108.2	32.8
Glycoside hydrolases	261	GH50	65.0	37.8	38.0	34.7	73.2	47.5	85.9	39.4
Glycoside hydrolases	262	GH50	142.5	66.0	154.2	69.9	118.6	49.1	131.0	66.4
Glycoside hydrolases	263	GH50	37.9	13.1	35.4	31.3	51.9	22.9	66.8	28.3
Glycoside hydrolases	264	GH50	10.2	8.5	6.9	5.9	10.3	10.0	12.0	8.5
Glycoside hydrolases	265	GH50	30.7	16.7	21.7	17.3	21.7	13.9	23.5	10.4
Glycoside hydrolases	266	GH50	64.8	42.6	50.1	38.3	41.2	21.3	55.2	33.9

Glycoside hydrolases	267	GH50	88.6	45.0	118.4	60.4	80.7	40.1	86.6	33.0
Glycoside hydrolases	268	GH51	28.6	11.0	25.2	17.1	33.1	18.2	34.3	15.3
Glycoside hydrolases	269	GH51	46.7	41.4	79.2	50.5	67.7	55.2	90.0	78.9
Glycoside hydrolases	270	GH51 (Top 10: row 8)	375.7	117.7	393.7	240.2	708.3	344.4	691.7	226.7
Glycoside hydrolases	271	GH51 (Top 10: row 10)	407.2	175.6	278.5	233.0	597.0	442.4	512.3	291.0
Glycoside hydrolases	272	GH53	32.8	19.9	21.9	13.1	24.3	14.2	24.6	13.5
Glycoside hydrolases	273	GH55	96.7	28.7	56.7	25.7	143.6	65.2	140.9	59.1
Glycoside hydrolases	274	GH6 (Top 10: row 5)	744.4	426.7	688.0	578.5	1033.3	455.3	1017.0	533.8
Glycoside hydrolases	275	GH6 (Top 10: row 4)	1396.2	389.3	1135.2	856.0	1037.8	1042.8	1034.5	1090.7
Glycoside hydrolases	276	GH65	94.4	49.3	83.3	41.1	84.1	49.3	74.2	29.2
Glycoside hydrolases	277	GH73	113.4	40.1	139.6	90.1	68.6	44.5	85.9	50.5
Glycoside hydrolases	278	GH77	35.1	21.6	29.6	34.0	51.4	32.9	65.6	52.0
Glycoside hydrolases	279	GH77	172.0	69.5	199.6	111.5	259.9	114.5	244.1	85.9
Glycoside hydrolases	280	GH78	26.3	16.8	19.4	12.6	26.0	13.1	22.9	10.8
Glycoside hydrolases	281	GH78	9.7	10.9	8.7	7.6	8.2	6.4	8.3	5.2
Glycoside hydrolases	282	GH78	6.5	7.2	2.8	2.6	29.0	32.5	16.0	25.4
Glycoside hydrolases	283	GH81 (Top 10: row 2)	14747.6	5244.6	18153.2	9815.9	18464.7	8891.4	21141.3	9641.5
Glycoside hydrolases	284	GH82	30.7	18.1	26.5	24.4	19.2	15.9	33.5	23.3
Glycoside hydrolases	285	GH82	68.7	22.5	70.9	43.4	90.3	48.0	83.3	35.2

Glycoside hydrolases	286	GH88	150.2	129.0	232.0	164.2	249.2	190.7	199.9	125.3
Glycoside hydrolases	287	GH9	43.9	19.2	39.0	18.7	70.2	36.4	66.1	35.7
Glycoside hydrolases	288	GH9	56.4	40.3	68.4	32.0	59.4	50.3	58.9	43.3
Glycoside hydrolases	289	GH9	36.6	28.1	26.5	15.4	41.8	52.2	28.9	14.4
Glycoside hydrolases	290	GH9	54.6	64.0	24.9	14.7	49.5	22.7	53.9	22.0
Glycoside hydrolases	291	GH92	18.6	16.7	16.8	12.3	14.1	8.0	14.9	8.3
Glycoside hydrolases	292	GH94	20.2	11.2	24.1	19.7	24.8	21.6	29.3	12.2
Glycoside hydrolases	293	GH94	49.7	15.3	54.5	28.9	46.3	28.1	53.5	23.8
Glycoside hydrolases	294	GH95	28.0	13.9	24.4	11.6	34.7	18.2	36.6	12.3
Glycoside hydrolases	295	GH95	355.3	470.2	134.0	45.5	108.6	111.2	92.2	67.8
Glycoside hydrolases	296	GH95	77.9	29.5	40.7	22.7	47.7	28.5	65.6	32.7
Glycoside hydrolases	297	GH95	8.0	3.5	5.9	4.5	4.6	3.0	5.6	3.5
Glycoside hydrolases	298	GH95	58.1	18.0	58.2	16.5	48.4	14.7	47.7	15.1
Glycoside hydrolases	299	GH95	13.0	9.5	17.1	17.0	24.8	15.2	26.5	17.8
Glycoside hydrolases	300	GH95	37.4	20.9	24.7	16.6	28.9	14.4	29.5	9.6
Glycoside hydrolases	301	GH95	13.4	7.0	15.1	9.3	17.8	11.0	23.1	10.7
Glycoside hydrolases	302	GH95	9.0	6.6	9.4	6.8	11.6	9.5	13.3	6.1
Glycoside hydrolases	303	GH95	70.0	43.4	69.1	46.8	71.0	42.2	62.6	31.0
Glycoside hydrolases	304	GH97	30.5	15.5	25.1	11.1	34.0	17.3	33.6	11.1

Glycoside hydrolases	305	GH97	37.1	13.0	32.6	17.8	48.8	11.5	54.4	25.9
Glycoside hydrolases	306	GH99	2.0	3.5	5.1	4.9	0.0	0.0	2.6	5.9
Motility & Chemotaxis	1	Chemotaxis protein CheD	11.4	8.8	13.6	8.2	16.5	8.1	12.1	11.1
Motility & Chemotaxis	2	Chemotaxis protein CheX	16.6	12.1	11.9	10.1	11.3	6.3	12.7	5.1
Motility & Chemotaxis	3	Chemotaxis protein methyltransferase CheR	8.1	5.1	6.5	4.8	6.2	6.4	7.5	9.7
Motility & Chemotaxis	4	Chemotaxis response regulator, CheB	10.0	6.4	9.9	6.6	11.4	7.3	12.5	5.7
Motility & Chemotaxis	5	Methyl-accepting chemotaxis protein	8.1	6.3	9.2	6.3	12.2	8.1	13.0	6.8
Motility & Chemotaxis	6	Methyl-accepting chemotaxis protein I	33.6	13.0	32.6	14.9	37.0	23.9	41.3	28.4
Motility & Chemotaxis	7	Flagellar basal-body rod modification protein FlgD	18.2	15.7	24.5	11.5	17.8	10.5	16.4	5.1
Motility & Chemotaxis	8	Flagellar basal-body rod protein FlgB	13.1	18.2	14.6	11.1	14.6	8.7	6.4	7.4
Motility & Chemotaxis	9	Flagellar basal-body rod protein FlgC	22.8	19.0	21.4	17.6	11.4	8.1	16.0	15.2
Motility & Chemotaxis	10	Flagellar basal-body rod protein FlgG	21.9	11.9	15.4	8.8	9.0	10.4	18.9	13.6
Motility & Chemotaxis	11	Flagellar biosynthesis protein FlhA	26.4	12.9	33.2	13.5	21.2	11.4	24.0	15.0
Motility & Chemotaxis	12	Flagellar biosynthesis protein FlhB	10.9	3.0	9.7	3.9	9.5	6.0	7.4	5.2
Motility & Chemotaxis	13	Flagellar biosynthesis protein FliP	11.7	7.4	7.8	5.4	11.0	6.8	8.5	4.3
Motility & Chemotaxis	14	Flagellar biosynthesis protein FliQ	6.1	5.8	8.6	6.5	10.7	7.2	6.4	8.3
Motility & Chemotaxis	15	Flagellar cap protein FliD	40.1	37.1	22.4	25.6	10.7	7.5	9.3	6.2
Motility & Chemotaxis	16	Flagellar hook protein FlgE	17.2	10.4	16.6	5.0	14.7	5.0	11.6	4.8
Motility & Chemotaxis	17	Flagellar L-ring protein FlgH	8.4	5.7	6.8	8.0	6.1	5.8	3.6	3.9

Motility & Chemotaxis	18	Flagellar motor rotation protein MotA	11.4	2.8	8.7	5.2	9.1	7.5	6.3	3.9
Motility & Chemotaxis	19	Flagellar motor rotation protein MotB	20.6	11.6	15.8	10.4	12.8	8.7	12.1	12.5
Motility & Chemotaxis	20	Flagellar M-ring protein FliF	44.8	40.1	59.6	47.3	20.4	15.7	23.4	14.2
Motility & Chemotaxis	21	Flagellar P-ring protein FlgI	14.2	7.1	11.9	6.1	10.2	6.4	11.8	4.1
Motility & Chemotaxis	22	Flagellum biosynthesis repressor protein FlbT	16.7	4.8	11.8	7.5	12.2	11.8	6.3	6.0
Motility & Chemotaxis	23	Flagellum-specific ATP synthase FliI	11.1	8.3	10.2	4.8	9.9	6.2	15.2	8.1
Sialic acid	1	Sialidase	37.0	13.5	36.6	12.3	22.6	10.2	23.8	6.1
Sialic acid	2	Predicted sialic acid transporter	14.3	8.9	12.6	6.4	13.8	8.6	16.9	6.4
Sialic acid	3	Predicted sialic acid transporter	25.4	13.4	21.6	8.7	21.3	9.5	21.4	8.0
Sialic acid	4	Sialic acid-specific 9-O-acetylesterase	40.3	14.1	33.9	9.1	35.7	15.6	37.1	15.5
Sialic acid	5	Sialic acid-specific 9-O-acetylesterase	52.6	20.5	51.8	30.7	38.0	23.8	43.7	25.7
Sialic acid	6	Sialic acid-specific 9-O-acetylesterase	19.9	11.0	20.1	8.8	20.0	14.4	26.1	15.5
Sigma factor	1	RNA polymerase sigma factor for flagellar operon	12.1	9.2	6.1	4.6	8.1	5.8	8.7	8.0
Sigma factor	2	RNA polymerase sigma factor RpoD	2280.4	1393.5	1821.5	712.5	3825.3	2186.1	3656.3	1617.6
Sigma factor	3	RNA polymerase sigma factor RpoD	136.5	106.8	72.2	27.2	124.1	79.7	154.3	122.7
Sigma factor	4	RNA polymerase sigma factor RpoE	994.3	392.8	1007.3	444.5	805.3	193.8	782.6	171.1
Sigma factor	5	RNA polymerase sigma factor RpoE	4369.9	1687.2	4321.5	1670.2	5916.6	2069.4	5920.2	2177.6
Sigma factor	6	RNA polymerase sigma factor RpoE	5276.2	4967.5	6295.6	3275.0	10800.5	4504.4	11492.5	5758.3
Sigma factor	7	RNA polymerase sigma factor RpoE	3181.2	4716.5	1886.4	1368.2	2207.7	1301.6	1944.8	1152.5
Sigma factor	8	RNA polymerase sigma-54 factor RpoN	55.6	29.6	47.6	24.1	43.5	20.0	48.4	24.0
Sigma factor	9	RNA polymerase sigma-54 factor RpoN	174.0	111.2	119.9	42.0	102.8	33.4	68.9	36.6
Sigma factor	10	RNA polymerase sigma-54 factor RpoN	267.2	181.1	64.8	38.8	41.9	17.0	48.7	14.2

Sigma factor	11	RNA polymerase sigma-70 factor, ECF subfamily	744.1	321.6	597.4	203.6	1034.6	610.8	924.5	609.6
Sulfatase	1	Sulfatase	37.9	22.7	46.1	39.9	50.6	36.4	52.4	33.8
Sulfatase	2	Sulfatase	7.6	7.0	9.5	6.6	15.4	10.0	12.8	6.2
Sulfatase	3	Sulfatase	38.9	25.3	65.2	38.1	50.4	43.8	55.4	33.4
Sulfatase	4	Sulfatase	97.0	37.5	101.5	43.0	63.2	31.1	77.6	29.9
Sulfatase	5	Sulfatase	31.6	15.9	29.9	15.3	29.5	17.6	31.6	14.6
Sulfatase	6	Sulfatase	9.6	6.2	8.5	5.9	13.1	7.0	16.1	10.5
Sulfatase	7	Sulfatase	25.2	16.5	19.6	13.4	28.8	15.4	24.4	15.9
Sulfatase	8	Sulfatase	32.8	19.9	21.9	13.1	24.3	14.2	24.6	13.5
Sulfatase	9	Sulfatase	29.9	9.7	28.9	7.9	38.5	23.0	42.7	20.7
Sulfatase	10	Sulfatase	14.2	7.5	11.1	1.4	10.1	4.3	14.1	5.4
Sulfatase	11	Sulfatase	84.9	32.5	39.7	18.5	34.1	17.2	41.9	13.9
Sulfatase	12	Sulfatase	19.3	13.3	20.4	7.6	21.3	10.0	21.8	10.8
Sulfatase	13	Sulfatase	4.4	4.2	3.4	2.1	5.3	3.2	8.3	1.2
Sulfatase	14	Sulfatase	8.9	5.2	7.2	5.6	12.1	7.0	13.6	6.7
Sulfatase	15	Sulfatase	16.2	13.4	18.1	10.9	17.1	10.1	23.4	17.2
Sulfatase	16	Sulfatase	10.3	5.8	7.1	5.8	8.4	4.8	13.5	5.0
Sulfatase	17	Sulfatase	6.7	4.7	6.4	4.7	9.0	4.8	6.9	4.5
Sulfatase	18	Sulfatase	43.8	24.2	42.4	21.5	30.1	13.0	39.7	12.1
Sulfatase	19	Sulfatase	44.7	34.2	34.2	25.1	39.7	24.9	48.4	22.6
Sulfatase	20	Sulfatase	15.4	12.1	5.7	4.9	11.2	5.2	13.4	8.8
Sulfatase	21	Sulfatase	50.9	31.3	34.9	21.2	28.7	10.1	27.6	14.0
Sulfatase	22	Sulfatase	16.5	15.2	10.9	10.7	19.6	19.2	20.3	14.3
Sulfatase	23	Sulfatase	40.2	22.3	30.4	13.9	40.6	18.0	38.4	14.5
Sulfatase	24	Sulfatase	9.8	7.5	5.4	5.2	6.7	7.2	4.5	3.2

Sulfatase	25	Sulfatase	24.4	31.2	26.1	18.2	24.9	13.6	22.5	9.2
Sulfatase	26	Sulfatase	38.1	38.3	62.4	80.1	136.6	192.9	109.4	134.6
Sulfatase	27	Sulfatase	5.6	3.8	3.9	2.0	8.1	12.1	4.7	5.7
Sulfatase	28	Sulfatase	19.9	16.5	14.3	9.1	21.0	16.5	20.5	15.6
Sulfatase	29	Sulfatase	25.3	22.0	22.2	17.0	31.6	15.4	30.7	13.5
Sulfatase	30	Sulfatase	74.2	41.2	52.8	37.7	61.8	36.4	67.6	27.2
Sulfatase	31	Sulfatase	19.8	13.1	23.6	15.5	26.7	11.4	28.4	12.4
Sulfatase	32	Sulfatase	88.8	50.1	118.6	83.5	134.3	65.8	145.3	68.7
Sulfatase	33	Sulfatase	10.7	5.8	9.3	5.5	14.2	8.2	12.5	6.8
Sulfatase	34	Sulfatase	17.3	13.1	18.3	15.1	23.9	12.3	28.5	7.3
Sulfatase	35	Sulfatase	48.2	25.3	57.0	24.7	41.6	21.7	46.2	16.2
Sulfatase	36	Sulfatase	215.1	136.4	176.5	126.3	116.6	59.5	130.3	49.6
Sulfatase	37	Sulfatase	35.9	17.1	46.4	11.1	91.3	59.4	74.9	31.3
Sulfatase	38	Sulfatase	9.1	5.7	7.4	4.4	6.7	4.6	10.9	11.9
Sulfatase	39	Sulfatase	29.4	7.5	28.6	17.8	36.8	23.1	33.5	21.3
Sulfatase	40	Sulfatase	9.7	4.7	6.6	4.0	8.6	4.1	9.3	8.0
Sulfatase	41	Sulfatase	20.5	11.7	17.9	9.7	27.3	20.4	26.2	16.7
Sulfatase	42	Sulfatase	88.7	49.7	74.3	74.7	121.0	80.5	126.0	77.9
Sulfatase	43	Sulfatase	130.9	79.9	160.8	101.4	192.2	116.0	212.1	124.9
Sulfatase	44	Sulfatase	19.9	11.9	18.7	16.5	18.7	10.0	23.7	20.6
Sulfatase	45	Sulfatase	15.4	6.7	12.0	8.8	16.5	7.6	18.6	7.9
Sulfatase	46	Sulfatase	26.0	14.3	27.9	18.6	25.0	13.3	26.1	16.8
Sulfatase	47	Sulfatase	51.9	36.5	38.8	29.3	52.8	37.9	53.7	37.4
Sulfatase	48	Sulfatase	51.8	33.2	44.4	26.2	35.9	21.1	51.8	26.3
Sulfatase	49	Sulfatase	43.4	20.6	42.3	13.2	22.7	11.1	32.3	17.9
Sulfatase	50	Sulfatase	11.9	4.4	10.9	8.7	12.9	4.9	15.6	5.9

Sulfatase	51	Sulfatase	72.9	55.6	76.7	50.4	68.4	30.6	62.5	24.7
Sulfatase	52	Sulfatase	24.2	13.6	27.9	14.9	29.8	16.9	28.5	12.8
Sulfatase	53	Sulfatase	40.1	36.1	17.3	13.6	20.0	19.0	14.8	13.9
Sulfatase	54	Sulfatase	10.1	4.0	10.9	5.9	11.2	7.5	12.1	7.4
Sulfatase	55	Sulfatase	14.7	9.0	13.7	9.2	17.1	9.9	18.1	10.7
Sulfatase	56	Sulfatase	21.5	11.5	19.0	12.2	27.5	12.7	22.8	10.4
Sulfatase	57	Sulfatase	28.5	16.3	27.0	18.4	21.2	10.6	21.1	9.7
Sulfatase	58	Sulfatase	50.3	23.2	25.6	17.3	32.4	20.8	30.4	22.4
Sulfatase	59	Sulfatase	20.5	6.9	13.0	5.7	13.9	5.9	17.3	5.5
Sulfatase	60	Sulfatase	13.6	8.6	10.2	6.6	12.8	6.9	17.4	9.1
Sulfatase	61	Sulfatase	46.1	20.0	41.3	16.0	41.1	19.1	44.0	11.1
Sulfatase	62	Sulfatase	43.1	22.7	41.3	26.6	54.6	36.8	65.8	42.8
Sulfatase	63	Sulfatase	8.4	7.1	6.3	5.8	11.9	4.8	12.5	10.8
Sulfur utilization	1	4Fe-4S ferredoxin iron-sulfur binding domain protein	103.5	38.5	59.5	25.7	134.7	43.8	146.7	39.7
Sulfur utilization	2	Adenylylsulfate kinase	46.6	24.4	31.0	15.0	38.2	24.5	42.3	17.1
Sulfur utilization	3	Cysteine desulfurase	34.1	15.4	27.8	14.2	29.2	8.3	21.8	8.1
Sulfur utilization	4	Cysteine desulfurase, SufS	282.1	90.9	180.1	98.2	193.3	77.5	206.3	49.8
Sulfur utilization	5	Cysteine synthase	648.4	699.2	2972.7	1480.3	4711.2	4560.9	3830.6	3573.8
Sulfur utilization	6	Ferredoxin	44.3	26.2	49.9	39.9	47.2	32.7	42.1	37.3
Sulfur utilization	7	Ferredoxin	110.4	50.1	93.5	77.3	116.2	59.7	139.2	50.6
Sulfur utilization	8	Iron-sulfur cluster assembly ATPase protein SufC	487.0	419.2	438.6	405.1	416.7	336.0	415.7	357.2
Sulfur utilization	9	Iron-sulfur cluster assembly protein SufB	518.3	474.3	340.5	360.6	340.9	271.2	417.9	321.1
Sulfur utilization	10	Iron-sulfur cluster assembly protein SufD	199.4	157.2	119.8	146.8	84.7	83.5	134.6	161.9

Sulfur utilization	11	Iron-sulfur cluster-binding protein	125.7	37.2	79.8	38.2	137.0	28.3	143.6	22.4
Sulfur utilization	12	Iron-sulfur cluster-binding protein	22.7	12.7	24.3	14.7	18.4	9.4	26.9	13.4
Sulfur utilization	13	Iron-sulfur cluster-binding protein	57.5	26.9	55.9	30.7	47.1	24.0	49.6	22.8
Sulfur utilization	14	PaaD-like protein (DUF59) involved in Fe-S cluster assembly	84.9	30.4	53.1	19.5	71.6	33.1	52.6	15.2
Sulfur utilization	15	PaaD-like protein (DUF59) involved in Fe-S cluster assembly	406.1	354.6	188.5	204.7	258.5	228.8	285.7	341.4
Sulfur utilization	16	Phosphoadenylyl-sulfate reductase [thioredoxin]	40.7	24.7	282.0	170.4	416.8	169.4	339.7	181.6
Sulfur utilization	17	Probable iron binding protein, SufA family	1101.3	1724.1	172.7	109.7	71.6	49.4	70.2	31.6
Sulfur utilization	18	Putative arylsulfatase regulatory protein	102.4	52.5	85.7	39.2	97.7	106.9	100.9	93.9
Sulfur utilization	19	Putative iron-sulfur cluster assembly scaffold protein for SUF system, SufE2	564.3	551.2	686.9	724.3	1377.2	995.7	1405.4	1027.3
Sulfur utilization	20	Sulfate adenylyltransferase / Adenylylsulfate kinase	79.1	66.8	46.2	17.4	37.6	23.1	48.2	28.2
Sulfur utilization	21	Sulfate adenylyltransferase subunit 1	142.6	222.0	37.5	9.5	46.1	33.1	47.0	23.4
Sulfur utilization	22	Sulfate adenylyltransferase subunit 2	103.2	121.6	67.0	9.3	107.9	55.5	80.3	35.3
Sulfur utilization	23	Sulfate permease	97.0	59.5	68.6	18.3	60.0	34.6	63.5	23.2
Sulfur utilization	24	Sulfate permease	72.0	21.5	61.1	23.6	47.6	20.3	50.5	15.7
Sulfur utilization	25	Sulfate permease, Trk-type	71.6	37.5	51.9	23.6	72.2	32.7	78.5	32.6
Sulfur utilization	26	Sulfide:quinone oxidoreductase, Type III	384.5	194.0	123.1	66.1	75.0	36.3	187.1	222.5
Sulfur utilization	27	Sulfite reductase [NADPH] flavoprotein alpha-component	76.9	35.1	46.3	21.4	65.2	22.5	57.5	11.9
Sulfur utilization	28	Sulfite reductase [NADPH] hemoprotein beta-component	46.9	18.7	81.3	14.1	144.1	106.0	121.0	92.0
Sulfur utilization	29	Sulfur acceptor protein SufE for iron-sulfur cluster assembly	24.2	17.1	21.9	8.3	26.9	10.8	23.4	10.7

Sulfur utilization	30	Sulfur carrier protein FdhD	27.8	22.7	29.4	14.8	21.1	7.0	41.0	14.1
TIV Pili	1	Type IV fimbrial assembly protein PilC	967.8	1425.2	303.8	126.8	159.9	75.9	192.7	98.8
TIV Pili	2	Type IV fimbrial assembly, ATPase PilB	197.7	133.0	56.1	33.6	120.1	46.3	125.1	31.4
TIV Pili	3	Type IV fimbrial assembly, ATPase PilB	286.8	269.6	92.5	33.7	106.7	70.0	104.1	56.6
TIV Pili	4	Type IV fimbrial assembly, ATPase PilB	35.9	19.3	33.6	23.3	26.9	16.0	30.0	15.1
TIV Pili	5	Type IV fimbrial assembly, ATPase PilB	91.8	96.3	21.1	10.1	43.5	15.2	56.0	22.2
Transport	1	ABC transporter, ATP-binding protein	47.6	16.6	33.7	23.8	36.1	18.6	38.5	25.7
Transport	2	ABC transporter, ATP-binding protein	89.1	40.8	39.6	16.9	47.8	26.8	59.6	29.4
Transport	3	ABC transporter, ATP-binding protein	50.4	44.9	40.8	43.5	103.4	133.0	113.1	155.2
Transport	4	ABC transporter, ATP-binding protein	77.6	37.9	50.6	33.0	67.3	30.5	75.4	26.7
Transport	5	ABC transporter, ATP-binding protein (cluster 9, phospholipid)	15.2	10.4	10.4	7.2	10.8	6.6	15.6	17.1
Transport	6	ABC transporter, ATP-binding protein (cluster 9, phospholipid)	21.9	9.9	20.0	12.3	20.4	9.3	24.7	19.4
Transport	7	ABC transporter, permease protein (cluster 9, phospholipid)	58.4	22.6	54.1	15.9	56.1	20.4	59.1	14.1
Transport	8	ABC transporter, permease protein (cluster 9, phospholipid)	136.3	146.4	36.2	12.3	45.6	20.2	52.0	26.9
Transport	9	ABC transporter, permease protein 1 (cluster 5, nickel/peptides/opines)	80.2	34.0	54.8	25.6	42.0	17.9	59.1	19.2
Transport	10	ABC transporter, permease protein 1 (cluster 5, nickel/peptides/opines)	20.5	11.8	16.6	8.2	12.5	7.2	13.8	7.2
Transport	11	ABC transporter, permease protein 2 (cluster 5, nickel/peptides/opines)	47.6	17.2	40.5	12.5	32.5	11.6	36.1	8.5
Transport	12	ABC transporter, RND-adapter-like protein	221.4	107.1	140.0	119.0	373.3	388.1	369.6	408.9
Transport	13	ABC transporter, substrate-binding protein (cluster 5, nickel/peptides/opines)	93.4	88.8	28.2	11.0	34.4	7.9	44.2	18.5

Transport	14	ABC transporter, substrate-binding protein (cluster 5, nickel/peptides/opines)	22.7	9.4	19.9	5.8	22.0	9.4	18.8	5.8
Transport	15	ABC-type antimicrobial peptide transport system, ATPase component	34.4	25.5	38.3	26.7	79.9	39.9	68.8	46.0
Transport	16	ABC-type antimicrobial peptide transport system, ATPase component	24.0	22.0	23.3	20.8	24.9	19.0	34.6	19.9
Transport	17	ABC-type antimicrobial peptide transport system, ATPase component	24.6	18.5	29.7	15.1	40.8	28.9	35.8	28.4
Transport	18	ABC-type antimicrobial peptide transport system, ATPase component	5.7	5.7	4.7	2.5	11.3	12.6	7.5	4.6
Transport	19	ABC-type antimicrobial peptide transport system, ATPase component	5.9	4.1	4.0	3.1	4.0	3.1	4.1	4.0
Transport	20	ABC-type antimicrobial peptide transport system, ATPase component	64.5	43.9	35.3	20.1	68.6	41.9	76.6	69.0
Transport	21	ABC-type antimicrobial peptide transport system, ATPase component	340.7	169.1	435.3	274.3	978.2	443.5	961.6	572.0
Transport	22	ABC-type antimicrobial peptide transport system, ATPase component	141.4	76.1	85.6	74.5	175.9	146.0	212.5	178.7
Transport	23	ABC-type antimicrobial peptide transport system, ATPase component	429.9	552.2	68.1	45.2	124.9	120.9	125.8	146.2
Transport	24	ABC-type antimicrobial peptide transport system, permease component	46.7	27.6	45.3	29.6	59.5	35.2	75.7	45.4
Transport	25	ABC-type antimicrobial peptide transport system, permease component	26.0	10.0	26.1	14.1	37.7	19.6	41.2	18.4
Transport	26	ABC-type antimicrobial peptide transport system, permease component	142.4	84.4	123.5	120.1	179.3	119.4	177.4	111.3
Transport	27	ABC-type antimicrobial peptide transport system, permease component	87.7	49.7	73.3	72.2	127.9	75.1	138.7	104.0
Transport	28	ABC-type Fe3+ transport system protein; Molybdenum transport protein, putative	157.5	210.0	29.3	10.3	36.2	26.2	31.4	10.2
Transport	29	ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component	21.4	7.2	13.7	6.8	30.2	13.8	23.8	12.1
Transport	30	Acetate permease ActP (cation/acetate symporter)	39.5	20.3	28.2	10.6	19.6	6.0	26.6	7.3
Transport	31	Acidobacterial duplicated orphan permease (function unknown)	185.2	155.3	153.8	83.3	225.5	99.9	249.6	124.4

Transport	32	Acidobacterial duplicated orphan permease (function unknown)	27.3	15.4	16.4	11.0	22.4	11.5	20.3	10.3
Transport	33	Acidobacterial duplicated orphan permease (function unknown)	50.7	23.8	15.9	11.2	18.6	11.7	24.0	8.8
Transport	34	Acidobacterial duplicated orphan permease (function unknown)	50.7	25.4	15.7	11.6	22.7	11.6	23.3	12.0
Transport	35	Ammonium transporter	345.4	359.3	66.3	26.5	102.3	55.6	110.6	58.4
Transport	36	Ammonium transporter	181.7	126.6	130.8	92.6	146.1	86.1	156.3	70.9
Transport	37	AttH component of AttEFGH ABC transport system	21.4	7.9	13.9	6.7	25.8	7.9	18.3	7.4
Transport	38	Biopolymer transport protein ExbD/TolR	123.2	99.3	35.8	18.5	88.5	51.2	75.5	27.4
Transport	39	Biopolymer transport protein ExbD/TolR	92.9	81.2	58.0	30.6	33.4	11.2	70.8	57.0
Transport	40	Biopolymer transport protein ExbD/TolR	393.3	208.6	437.7	116.0	264.3	103.3	309.9	107.6
Transport	41	Biopolymer transport protein ExbD/TolR	84.6	105.0	28.8	12.6	15.3	4.2	23.3	11.0
Transport	42	Biopolymer transport protein ExbD/TolR	35.0	21.1	32.9	22.4	28.0	27.7	35.1	19.3
Transport	43	Biopolymer transport protein ExbD/TolR	426.2	289.4	133.4	31.3	218.6	113.6	210.5	142.1
Transport	44	Biopolymer transport protein ExbD/TolR	1691.2	994.9	1527.9	1056.6	563.7	112.9	621.7	147.6
Transport	45	Cation transport protein	73.0	28.9	55.4	46.5	86.1	45.6	103.9	55.1
Transport	46	COG0385 sodium-dependent transporter	63.1	20.9	65.5	33.7	66.6	40.5	59.7	33.6
Transport	47	D-glycerate transporter (predicted)	38.4	23.1	48.3	28.0	45.3	20.3	45.1	23.5
Transport	48	D-glycerate transporter (predicted)	62.5	47.4	67.8	50.8	75.7	58.2	77.5	62.6
Transport	49	D-glycerate transporter (predicted)	50.3	45.9	36.6	24.3	23.7	9.4	28.0	9.7
Transport	50	Dipeptide-binding ABC transporter, periplasmic substrate-binding component	22.5	15.1	17.0	11.1	11.3	6.5	14.6	6.7
Transport	51	D-xylose proton-symporter XylE	31.0	24.8	18.0	7.5	9.7	6.4	10.0	11.0
Transport	52	D-xylose proton-symporter XylE	13.6	7.8	17.5	7.8	22.1	9.2	19.0	14.4

Transport	53	D-xylose proton-symporter XylE	9.6	3.2	8.8	1.8	11.4	4.1	9.4	4.7
Transport	54	Efflux ABC transporter, ATP-binding protein	139.8	63.0	104.5	48.2	681.3	566.7	605.2	610.9
Transport	55	Efflux ABC transporter, ATP-binding protein	170.0	72.3	118.9	45.9	108.4	62.1	132.6	84.7
Transport	56	Efflux ABC transporter, permease/ATP-binding protein	70.5	32.4	27.3	7.8	29.8	14.1	33.9	14.7
Transport	57	Efflux transport system, outer membrane factor (OMF) lipoprotein	9.6	7.6	8.1	5.6	9.3	8.6	9.1	6.4
Transport	58	Ferrous iron transport protein B	220.2	76.4	174.1	62.7	248.9	117.6	249.7	121.0
Transport	59	Ferrous iron transport protein B	158.0	106.2	126.8	61.8	217.0	122.9	236.6	154.0
Transport	60	Ferrous iron transport protein B	82.1	20.6	60.6	23.8	90.4	41.2	119.1	55.0
Transport	61	Fluoride ion transporter CrcB	28.2	16.2	31.2	23.1	32.1	17.0	49.3	30.4
Transport	62	Fluoride ion transporter CrcB	28.7	13.3	20.0	13.0	17.0	17.8	16.9	19.9
Transport	63	Gliding motility-associated ABC transporter ATP-binding protein GldA	172.3	74.1	95.3	53.8	223.2	163.9	234.6	149.8
Transport	64	Gliding motility-associated ABC transporter permease protein GldF	54.5	29.7	36.1	9.1	91.4	43.0	68.8	30.5
Transport	65	Hemin ABC transporter, permease protein	503.7	238.1	617.2	281.9	308.7	134.5	303.3	123.5
Transport	66	Heterodimeric efflux ABC transporter, permease/ATP-binding subunit 1	441.9	236.5	448.9	280.4	216.0	153.8	235.6	198.1
Transport	67	Heterodimeric efflux ABC transporter, permease/ATP-binding subunit 2	244.5	122.8	243.1	231.7	140.9	82.0	182.0	162.3
Transport	68	Hydroxymethylpyrimidine ABC transporter, transmembrane component	25.9	12.1	14.6	9.6	34.3	15.3	30.0	11.0
Transport	69	Lipopolysaccharide ABC transporter, ATP-binding protein LptB	67.3	34.3	61.4	35.1	65.6	22.9	64.5	11.0
Transport	70	Long-chain fatty acid transport protein	217.1	138.7	206.4	124.9	92.8	32.5	92.5	41.9
Transport	71	L-rhamnose-proton symporter	84.1	21.6	105.5	30.8	53.9	24.7	61.4	25.8
Transport	72	L-rhamnose-proton symporter	23.9	19.3	10.8	8.6	35.9	35.9	39.7	43.6
Transport	73	Lysine exporter protein LysE/YggA	107.3	154.5	51.6	53.9	53.6	60.5	38.6	29.4
Transport	74	Lysophospholipid transporter LplT	37.1	15.1	46.5	20.9	30.1	15.7	35.2	15.7

Transport	75	Magnesium and cobalt efflux protein CorC	37.5	25.6	35.5	26.3	32.0	13.6	44.5	17.8
Transport	76	Mg/Co/Ni transporter MgtE, CBS domain-containing	178.3	86.5	195.2	45.0	152.5	72.1	149.0	60.1
Transport	77	MotA/TolQ/ExbB proton channel family protein	134.0	39.0	134.0	39.5	79.0	37.6	105.5	53.2
Transport	78	MotA/TolQ/ExbB proton channel family protein	79.1	86.2	38.6	14.9	44.7	23.8	49.1	29.9
Transport	79	MotA/TolQ/ExbB proton channel family protein	837.0	404.8	442.1	222.0	538.8	288.6	600.2	416.4
Transport	80	MotA/TolQ/ExbB proton channel family protein	56.7	28.7	29.5	16.4	22.8	24.9	31.8	10.4
Transport	81	MotA/TolQ/ExbB proton channel family protein	23.2	17.2	12.9	8.6	14.7	12.8	17.8	9.3
Transport	82	Multi antimicrobial extrusion protein	11.4	7.4	11.6	9.8	19.1	10.2	21.8	3.6
Transport	83	Multi antimicrobial extrusion protein	27.6	18.3	25.0	22.8	32.9	21.7	37.1	38.1
Transport	84	Na(+) H(+) antiporter subunit A	12.3	4.5	12.3	5.4	16.1	8.1	16.4	6.5
Transport	85	Na(+) H(+) antiporter subunit B	7.1	6.3	7.0	5.3	15.6	9.5	8.7	5.3
Transport	86	Na(+) H(+) antiporter subunit C	9.7	7.8	8.1	5.9	10.8	7.9	12.0	10.5
Transport	87	Na(+) H(+) antiporter subunit D	16.5	8.6	9.6	5.4	16.0	10.0	19.9	3.9
Transport	88	Na(+) H(+) antiporter subunit G	13.6	6.2	8.0	10.7	17.5	16.7	6.9	9.5
Transport	89	Na+/H+ antiporter	111.2	83.4	82.9	49.5	100.8	40.7	119.0	43.3
Transport	90	Na+/solute symporter	137.4	102.9	117.2	98.1	49.0	34.8	59.7	53.9
Transport	91	N-acetylglucosamine related transporter, NagX	32.6	22.9	31.4	13.4	40.2	15.2	53.5	27.1
Transport	92	Nucleoside permease NupC	70.0	30.9	67.6	43.7	116.8	70.6	121.8	61.8
Transport	93	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	55.4	30.2	30.9	23.5	61.1	23.5	59.2	17.2
Transport	94	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	60.2	18.2	46.7	17.8	54.4	21.5	52.9	12.2

Transport	95	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	31.2	15.3	19.5	9.1	24.5	14.4	26.9	11.7
Transport	96	Oligopeptide transport ATP-binding protein OppD	56.2	34.3	18.9	9.4	28.8	7.8	24.2	6.7
Transport	97	Oligopeptide transport ATP-binding protein OppD	103.1	57.2	64.5	70.6	107.1	59.1	118.3	61.8
Transport	98	Oligopeptide transport ATP-binding protein OppD	18.2	13.3	11.4	10.4	10.5	5.7	11.1	10.8
Transport	99	Oligopeptide transport ATP-binding protein OppF	31.2	18.3	15.3	4.5	16.8	3.2	23.0	16.1
Transport	100	Oligopeptide transport ATP-binding protein OppF	17.7	10.3	15.6	7.3	26.5	17.9	19.5	6.6
Transport	101	Oligopeptide transport ATP-binding protein OppF	148.0	84.5	149.0	131.4	291.7	360.6	257.2	257.4
Transport	102	Oligopeptide transport ATP-binding protein OppF	17.5	12.5	11.3	3.7	15.2	17.0	9.5	8.0
Transport	103	Oligopeptide transport system permease protein OppB	128.9	90.8	70.8	45.1	136.1	102.9	125.2	67.2
Transport	104	Oligopeptide transport system permease protein OppB	15.4	8.9	11.6	4.7	13.5	15.5	11.6	8.3
Transport	105	Oligopeptide transport system permease protein OppC	102.7	91.4	71.0	77.3	134.0	75.8	135.9	66.6
Transport	106	Pedicted L-rhamnose permease, NCS1 Family	50.7	28.1	31.4	29.0	33.2	16.7	41.1	10.9
Transport	107	Permease of the drug/metabolite transporter (DMT) superfamily	134.2	78.6	79.7	44.3	74.7	36.5	73.9	42.7
Transport	108	Permease of the drug/metabolite transporter (DMT) superfamily	21.6	11.5	19.5	6.4	23.7	11.8	21.2	10.0
Transport	109	Phosphate ABC transporter, periplasmic phosphate-binding protein PstS	756.3	567.8	113.1	62.3	249.2	158.1	249.3	158.8
Transport	110	Phosphate transport ATP-binding protein PstB	254.4	146.4	200.0	184.9	353.4	223.3	376.7	185.8
Transport	111	Phosphate transport system permease protein PstA	269.6	170.7	183.4	181.9	456.9	269.9	517.3	307.9
Transport	112	Phosphate transport system permease protein PstC	180.1	86.7	110.0	111.1	289.7	173.5	314.6	204.6

Transport	113	Phosphate transport system regulatory protein PhoU	302.0	145.5	265.0	226.3	377.0	163.0	373.5	177.7
Transport	114	Potassium efflux system KefA protein	76.6	39.9	50.7	17.0	46.2	13.1	56.1	19.5
Transport	115	Potassium efflux system KefA protein	146.9	99.7	75.1	27.4	60.5	33.8	76.4	34.5
Transport	116	Potassium efflux system KefA protein	84.8	26.7	77.0	29.7	66.0	37.7	70.3	30.0
Transport	117	Predicted maltose transporter MalT	51.7	33.0	74.8	108.4	83.9	70.6	98.3	77.1
Transport	118	Predicted sialic acid transporter	14.3	8.9	12.6	6.4	13.8	8.6	16.9	6.4
Transport	119	Predicted sialic acid transporter	25.4	13.4	21.6	8.7	21.3	9.5	21.4	8.0
Transport	120	Predicted sodium-dependent galactose transporter	39.0	18.4	33.3	10.2	28.6	5.9	29.8	7.1
Transport	121	Predicted sodium-dependent galactose transporter	48.0	23.9	46.0	33.0	48.1	23.4	58.6	32.7
Transport	122	Predicted sodium-dependent mannose transporter	19.7	11.4	20.8	10.8	22.2	9.7	22.9	11.5
Transport	123	Protein export cytoplasm protein SecA	19.3	14.9	27.9	31.7	29.0	48.3	19.9	19.6
Transport	124	putative divalent heavy-metal cations transporter	86.4	36.5	74.0	35.7	57.7	12.5	53.9	11.1
Transport	125	Putative nucleoside transporter yegT	65.0	18.3	48.7	20.5	43.0	18.5	40.8	20.8
Transport	126	Putative transport protein	61.8	28.7	59.5	29.8	43.1	39.9	55.7	31.8
Transport	127	Putative transport protein	53.3	15.2	44.6	29.5	65.3	35.2	67.7	46.8
Transport	128	Ribose ABC transport system, ATP-binding protein RbsA	45.2	21.6	34.3	28.9	26.5	22.0	30.7	16.0
Transport	129	Ribose ABC transport system, periplasmic ribose-binding protein RbsB	100.1	30.3	89.4	32.2	60.4	29.4	70.0	31.4
Transport	130	Ribose ABC transport system, permease protein RbsC	25.6	21.7	20.0	19.4	18.6	15.1	22.6	16.7
Transport	131	RND efflux system, inner membrane transporter	47.5	15.3	30.5	13.1	55.6	20.4	73.6	28.9
Transport	132	Sodium/glucose cotransporter 1	37.2	22.7	29.3	30.7	35.8	17.8	42.5	19.6
Transport	133	Sodium/sugar cotransporter	34.8	16.3	22.2	11.3	16.4	12.9	25.2	13.2
Transport	134	Sodium-Choline Symporter	20.5	14.8	12.2	8.1	13.1	7.9	16.5	6.1

Transport	135	Sodium-dependent anion transporter family	63.3	40.0	48.6	22.3	40.9	20.0	48.4	19.4
Transport	136	Sodium-dependent transporter	24.0	19.2	18.3	13.1	23.3	14.2	32.9	18.8
Transport	137	Sodium-dependent transporter, SNF family	65.7	26.5	40.1	12.0	60.4	14.3	78.4	37.6
Transport	138	sodium-solute symporter, putative	177.9	110.4	200.2	150.5	169.6	82.5	164.2	95.5
Transport	139	sodium-solute symporter, putative	58.2	28.3	59.6	29.3	72.1	27.3	80.0	41.6
Transport	140	sodium-solute symporter, putative	13.5	11.9	12.4	8.0	9.0	6.7	10.7	6.3
Transport	141	sodium-solute symporter, putative	31.1	9.1	29.0	12.4	25.9	22.3	21.5	16.5
Transport	142	sugar and carbohydrate transporters	33.5	3.3	25.6	14.4	24.5	12.4	27.4	18.5
Transport	143	TonB-dependent receptor	302.6	162.6	289.2	178.5	238.9	185.4	280.6	187.7
Transport	144	TonB-dependent receptor	24.1	16.7	25.5	19.8	37.9	25.0	38.8	24.1
Transport	145	TonB-dependent receptor	9.8	6.5	6.7	6.2	10.0	7.5	12.8	8.7
Transport	146	TonB-dependent receptor	78.8	36.7	78.8	42.3	63.3	27.0	70.3	20.4
Transport	147	TonB-dependent receptor plug	791.4	293.1	265.2	173.6	127.7	80.2	175.0	169.0
Transport	148	Transporter, sodium/sulfate symporter family	30.7	14.9	33.0	11.3	21.1	10.2	27.9	8.9
Transport	149	TRAP-type C4-dicarboxylate transport system, large permease component	46.8	32.3	21.1	12.7	24.8	12.6	31.0	16.8
Transport	150	TRAP-type C4-dicarboxylate transport system, periplasmic component	45.4	37.3	18.7	9.9	32.8	15.2	33.8	18.1
Transport	151	Trk potassium uptake system protein TrkA	85.3	35.1	54.7	19.9	65.2	25.3	65.0	14.3
Transport	152	Trk potassium uptake system protein TrkH	24.5	9.8	22.1	10.0	30.8	10.5	38.6	6.7
Transport	153	Twin-arginine translocation protein TatA	336.8	161.3	217.3	55.7	219.3	94.3	187.3	76.0
Transport	154	Twin-arginine translocation protein TatC	53.4	17.3	49.5	15.9	34.1	14.6	33.0	23.0
Transport	155	Uncharacterized amino acid permease, GabP family	41.5	19.4	45.4	17.4	38.4	17.2	47.6	11.3
Transport	156	Uncharacterized MFS-type transporter	177.7	71.7	185.5	89.7	195.2	85.2	190.9	66.9
Transport	157	Uncharacterized MFS-type transporter	30.9	10.2	22.5	9.9	28.4	15.7	35.3	11.8

Transport	158	Uncharacterized MFS-type transporter	46.1	47.0	25.3	18.5	12.0	5.4	18.8	11.6
Transport	159	Uncharacterized sugar:proton symporter	38.3	19.3	29.8	9.8	24.9	14.5	35.0	14.6
Transport	160	Uncharacterized sugar:proton symporter	48.0	28.3	41.7	26.9	30.7	20.8	34.0	15.6
Transport	161	Uncharacterized sugar:proton symporter	21.2	13.1	32.4	15.1	27.4	10.4	32.0	12.4
Transport	162	Vitamin B12 ABC transporter, substrate-binding protein BtuF	62.0	23.6	73.6	44.6	59.9	35.8	50.5	29.0
Transport	163	Xyloside transporter XynT	16.6	18.0	12.6	8.3	9.0	8.3	13.2	13.4

Literature Cited

1. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey H V, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A 110:3229–36.

2. Douglas AE. 2009. The microbial dimension in insect nutritional ecology. Funct Ecol 23:38–47.

3. Wilson ACC, Duncan RP. 2015. Signatures of host / symbiont genome coevolution in insect nutritional endosymbioses. Proc Natl Acad Sci 112:10255–10261.

4. Harmer TL, Rotjan RD, Nussbaumer AD, Bright M, Ng AW, Dechaine EG, Cavanaugh CM. 2008. Free-Living Tube Worm Endosymbionts Found at Deep-Sea Vents. Appl Environ Microbiol 74:3895–3898.

5. Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. 2017. The evolution of host-symbiont dependence. Nat Commun 8:15973.

6. Engl T, Kroiss J, Kai M, Nechitaylo TY. 2018. Evolutionary stability of antibiotic protection in a defensive symbiosis. Proc Natl Acad Sci 115:E2020–E2029.

7. Stecher B, Hardt W. 2008. The role of microbiota in infectious disease. Trends Microbiol 16:107–114.

8. Weingarden A, González A, Vázquez-baeza Y, Weiss S, Humphry G, Berg-lyons D, Knights D, Unno T, Bobr A, Kang J, Khoruts A, Knight R, Sadowsky MJ. 2015. Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent *Clostridium difficile* infection. Microbiome 3:1–8.

9. Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Bakker PAHM. 2014. Induced Systemic Resistance by Beneficial Microbes. Annu Rev Phytopathol 52:347–375.

10. Pan X, Zhou G, Wu J, Bian G, Lu P, Raikhel AS, Xi Z. 2012. *Wolbachia* induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito *Aedes aegypti*. Proc Natl Acad Sci 109:E23–E31.

11. Round JL, Mazmanian SK. 2009. The gut microbiota shapes intestinal immune responses during health and disease. Nat Rev Immunol 9:313–323.

12. Bewley CA. 1996. Two classes of metabolites from *Theonella swinhoei* are localized in distinct populations of bacterial symbionts. Experientia 52:716–722.

13. Bubb MR, Spector I, Bershadsky AD, Korn ED. 1995. Swinholide A is a microfilament disrupting marine toxin that stabilizes actin dimers and severs actin filaments. J Biol Chem 270:3463–3466.

14. Schmidt EW, Bewley CA, Faulkner DJ. 1998. Theopalauamide, a Bicyclic Glycopeptide from Filamentous Bacterial Symbionts of the Lithistid Sponge *Theonella swinhoei* from

Palau and Mozambique. J Org Chem 63:1254–1258.

15. Schmidt EW, Obraztsova AY, Davidson SK, Faulkner DJ, Haygood MG. 2000. Identification of the antifungal peptide-containing symbiont of the marine sponge *Theonella swinhoei* as a novel deltaproteobacterium, "*Candidatus* Entotheonella palauensis'''." Mar Biol 136:969–977.

16. Brucker RM, Harris RN, Schwantes CR, Gallaher TN, Flaherty DC, Lam BA, Minbiole KPC. 2008. Amphibian Chemical Defense: Antifungal Metabolites of the Microsymbiont *Janthinobacterium lividum* on the Salamander *Plethodon cinereus*. J Chem Ecol 34:1422–1429.

17. Barke J, Seipke RF, Grüschow S, Heavens D, Drou N, Bibb MJ, Goss RJM, Yu DW, Hutchings MI. 2010. A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant *Acromyrmex octospinosus*. BMC Biol 8:1–10.

18. Oh DC, Poulsen M, Currie CR, Clardy J. 2009. Dentigerumycin: A bacterial mediator of an ant-fungus symbiosis. Nat Chem Biol 5:391–393.

19. Hamdoun A, Epel D. 2007. Embryo stability and vulnerability in an always changing world. Proc Natl Acad Sci 104:1745–50.

20. Byrne M. 2011. Impact of ocean warming and ocean acidification on marine invertebrate life history stages: vulnerabilities and potential for persistence in a changing ocean. Oceanogr Mar Biol An Annu Rev 49:1–42.

21. Mendes R, Garbeva P, Raaijmakers JM. 2013. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. FEMS Microbiol Rev 37:634–663.

22. Maranger R, Bird DF. 1995. Viral abundance in aquatic systems: a comparison between marine and fresh waters. Mar Ecol Prog Ser 121:217–226.

23. Zobell C, Allen E. 1935. The significance of marine bacteria in the fouling of submerged surfaces. J Bacteriol 29:239–251.

24. Cohen CS, Strathmann RR. 1996. Embryos at the edge of tolerance: Effects of environment and structure of egg masses on supply of oxygen to embryos. Biol Bull 190:8–15.

25. Fernández M, Bock C, Pörtner HO. 2000. The cost of being a caring mother: The ignored factor in the reproduction of marine invertebrates. Ecol Lett 3:487–494.

26. Rocha F. 2001. A review of reproductive strategies in cephalopods. Biol Rev 76:291–304.

27. Jonchère V, Réhault-godbert S, Hennequet-antier C, Cabau C, Sibut V, Cogburn LA, Nys Y, Gautron J. 2010. Gene expression profiling to identify eggshell proteins involved in physical defense of the chicken egg. BMC Genomics 11:1–19.

28. Pechenik JA. 1979. Role of encapsulation in invertebrate life histories. Am Nat 114:859–870.

29. Fraune S, Augustin R, Anton-Erxleben F, Wittlieb J, Gelhaus C, Klimovich VB,

Samoilovich MP, Bosch TCG. 2010. In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. Proc Natl Acad Sci 107:18067–72.

30. Baron OL, van West P, Industri B, Ponchet M, Dubreuil G, Gourbal B, Reichhart JM, Coustau C. 2013. Parental Transfer of the Antimicrobial Protein LBP/BPI Protects *Biomphalaria glabrata* Eggs against Oomycete Infections. PLoS Pathog 9:1–10.

31. Flórez L V., Biedermann PHW, Engl T, Kaltenpoth M. 2015. Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. Nat Prod Rep 32:879–1156.

32. Stock SP, Blair HG. 2008. Entomopathogenic nematodes and their bacterial symbionts: the inside out of a mutualistic association. Symbiosis 46:65–75.

33. Crawford JM, Portmann C, Zhang X, Roeffaers MBJ, Clardy J. 2012. Small molecule perimeter defense in entomopathogenic bacteria. Proc Natl Acad Sci 109:10821–10826.

34. Chen G, Zhang Y, Li J, Dunphy GB, Punja ZK, Webster JM. 1996. Chitinase Activity of *Xenorhabdus* and *Photorhabdus* Species, Bacterial Associates of Entomopathogenic Nematodes. J Invertebr Pathol 68:101–108.

35. Isaacson PJ, Webster JM. 2002. Antimicrobial activity of *Xenorhabdus* sp. RIO (Enterobacteriaceae), symbiont of the entomopathogenic nematode, *Steinernema riobrave* (Rhabditida: Steinernematidae). J Invertebr Pathol 79:146–153.

36. Li J, Hu K, Webster JM. 1998. Antibiotics from *Xenorhabdus* spp. and *Photorhabdus* spp. (enterobacteriaceae). Chem Heterocycl Compd 34:1331–1339.

37. Zhou X, Kaya HK, Heungens K, Goodrich-Blair H. 2002. Response of ants to a deterrent factor(s) produced by the symbiotic bacteria of entomopathogenic nematodes. Appl Environ Microbiol 68:6202–6209.

38. Kellner RLL. 2002. Molecular identification of an endosymbiotic bacterium associated with pederin biosynthesis in *Paederus sabaeus* (Coleoptera: Staphylinidae). Insect Biochem Mol Biol 32:389–395.

39. Kellner RLL, Dettner K. 1996. Differential efficacy of toxic pederin in deterring potential arthropod predators of *Paederus* (Coleoptera: Staphylinidae) offspring. Oecologia 107:293–300.

40. Flórez L V., Scherlach K, Miller IJ, Rodrigues A, Kwan JC, Hertweck C, Kaltenpoth M. 2018. An antifungal polyketide associated with horizontally acquired genes supports symbiont-mediated defense in Lagria villosa beetles. Nat Commun 9:2478.

41. Flórez L V, Scherlach K, Gaube P, Ross C, Sitte E, Hermes C, Rodrigues A, Hertweck C, Kaltenpoth M. 2017. Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect defensive mutualism. Nat Commun 8:15172.

42. Wang L, Feng Y, Tian J, Xiang M, Sun J, Ding J, Yin W-B, Stadler M, Che Y, Liu X. 2015. Farming of a defensive fungal mutualist by an attelabid weevil. ISME J 9:1793–801.

43. Gilliam M. 1997. Identication and roles of non-pathogenic microflora associated with honey bees. FEMS Microbiol Lett 155:1–10.

44. Forsgren E, Olofsson TC, Vásquez A, Fries I. 2010. Novel lactic acid bacteria inhibiting *Paenibacillus larvae* in honey bee larvae. Apidologie 41:99–108.

45. Kaltenpoth M, Goettler W, Dale C, Stubblefield JW, Herzner G, Roeser-mueller K, Strohm E. 2006. '*Candidatus* Streptomyces philanthi', an endosymbiotic streptomycete in the antennae of *Philanthus* digger wasps. Int J Syst Evol Microbiol 56:1403–1411.

46. Kroiss J, Kaltenpoth M, Schneider B, Schwinger M-G, Hertweck C, Maddula RK, Strohm E, Svatos A. 2010. Symbiotic Streptomycetes provide antibiotic combination prophylaxis for wasp offspring. Nat Chem Biol 6:261–263.

47. Kaltenpoth M, Gottler W, Herzner G, Strohm E. 2005. Symbiotic bacteria protect wasp larvae from fungal infestation. Curr Biol 15:475–479.

48. Koehler S, Doubský J, Kaltenpoth M. 2013. Dynamics of symbiont-mediated antibiotic production reveal efficient long-term protection for beewolf offspring. Front Zool 10:3.

49. Lam K, Thu K, Tsang M, Moore M, Gries G. 2009. Bacteria on housefly eggs, *Musca domestica*, suppress fungal growth in chicken manure through nutrient depletion or antifungal metabolites. Naturwissenschaften 96:1127–1132.

50. Soler JJ, Martin-Vivaldi M, Peralta-Sanchez JM, Arco L, Juarez-Garcia-Pelayo N. 2014. Hoopoes color their eggs with antimicrobial uropygial secretions. Naturwissenschaften 101:697–705.

51. Martín-Vivaldi M, Peña A, Peralta-Sánchez JM, Sánchez L, Ananou S, Ruiz-Rodríguez M, Soler JJ. 2010. Antimicrobial chemicals in hoopoe preen secretions are produced by symbiotic bacteria. Proc Biol Sci 277:123–130.

52. Ruiz-Rodríguez M, Martínez-Bueno M, Martín-Vivaldi M, Valdivia E, Soler JJ. 2013. Bacteriocins with a broader antimicrobial spectrum prevail in enterococcal symbionts isolated from the hoopoe's uropygial gland. FEMS Microbiol Ecol 85:495–502.

53. Martin-Vivaldi M, Soler JJ, Peralta-Sanchez JM, Arco L, Martin-Platero AM, Martinez-Bueno M, Ruiz-Rodriguez M, Valdivia E. 2014. Special structures of hoopoe eggshells enhance the adhesion of symbiont-carrying uropygial secretion that increase hatching success. J Anim Ecol 83:1289–1301.

54. Gil-Turnes M, Hay M, Fenical W. 1989. Symbiotic marine bacteria chemically defend crustacean embryos from a pathogenic fungus. Science (80-) 246:116–118.

55. Gil-Turnes M, Fenical W. 1992. Embryos of *Homarus americanus* are protected by epibiotic bacteria. Biol Bull 182:105–108.

56. Haygood MG, Davidson SK. 1997. Small-Subunit rRNA Genes and In Situ Hybridization with Oligonucleotides Specific for the Bacterial Symbionts in the Larvae of the Bryozoan *Bugula neritina* and Proposal of "*Candidatus* Endobugula sertula." Appl Environ Microbiol 63:4612–4616.

57. Lim GE, Haygood MG. 2004. "*Candidatus* Endobugula glebosa," a specific bacterial symbiont of the marine bryozoan *Bugula simplex*. Appl Environ Microbiol 70:4921–4929.

58. Davidson SK, Allen SW, Lim GE, Anderson CM, Haygood MG. 2001. Evidence for the Biosynthesis of Bryostatins by the Bacterial Symbiont "*Candidatus* Endobugula sertula" of the Bryozoan *Bugula neritina*. Appl Environ Microbiol 67:4531–4537.

59. Mutter R, Wills M. 2000. Chemistry and Clinical Biology of the Bryostatins. Bioorg Med Chem 8:1841–1860.

60. Lopanik N, Lindquist N, Targett N. 2004. Potent cytotoxins produced by a microbial symbiont protect host larvae from predation. Oecologia 139:131–139.

61. Lindquist N, Hay ME, Fenical W. 1992. Defense of ascidians and their conspicuous larvae: adult vs. larval chemical defenses. Ecol Monogr 62:547–568.

62. Sings HL, Rinehart KL. 1996. Compounds produced from potential tunicate-blue-green algal symbiosis: a review. Jounral Ind Microbiol Biotechnol 17:385–396.

63. Xu Y, Kersten RD, Nam SJ, Lu L, Al-Suwailem AM, Zheng H, Fenical W, Dorrestein PC, Moore BS, Qian PY. 2012. Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. J Am Chem Soc 134:8625–8632.

64. Tianero MDB, Kwan JC, Wyche TP, Presson AP, Koch M, Barrows LR, Bugni TS, Schmidt EW. 2015. Species specificity of symbiosis and secondary metabolism in ascidians. ISME J 9:615–628.

65. Buchner P. 1965. Symbiosis in luminous animals, p. 543–571. *In* Endosymbiosis of Animals with Plant Microorganisms.

66. Pierantoni U. 1918. Organi luminosi, organi simbiotici e glandola nidamentale accessoria nei Cefalopodi. Boll della Soc dei Nat Napoli 30:30–36.

67. Bloodgood RA. 1977. The squid accessory nidamental gland: Ultrastructure and association with bacteria. Tissue Cell 9:197–208.

68. Barbieri E, Paster BJ, Hughes D, Zurek L, Moser DP, Teske A, Sogin ML. 2001. Phylogenetic characterization of epibiotic bacteria in the accessory nidamental gland and egg capsules of the squid *Loligo pealei* (Cephalopoda: Loliginidae). Environ Microbiol 3:151–167.

69. Collins AJ, LaBarre BA, Wong Won BS, Shah M V., Heng S, Choudhury MH, Haydar SA, Santiago J, Nyholm S V. 2012. Diversity and partitioning of bacterial populations within the accessory nidamental gland of the squid *Euprymna scolopes*. Appl Environ Microbiol 78:4200–4208.

70. Biggs J, Epel D. 1991. Egg Capsule Sheath of *Loligo opalescens* Berry: Structure and Association with Bacteria. J Exp Zool 259:263–267.

71. Pichon D, Gaia V, Norman MD, Boucher-Rodoni R. 2005. Phylogenetic diversity of epibiotic bacteria in the accessory nidamental glands of squids (Cephalopoda: Loliginidae and Idiosepiidae). Mar Biol 147:1323–1332.

72. Kerwin AH, Nyholm S V. 2017. Symbiotic bacteria associated with a bobtail squid reproductive system are detectable in the environment, and stable in the host and in eggs throughout development. Environ Microbiol 19:1463–1475.

73. Grigioni S, Boucher-Rodoni R, Demarta A, M. T, Peduzzi R. 2000. Phylogenetic characterisation of bacterial symbionts in the accessory nidamental glands of the sepioid *Sepia officinalis* (Cephalopoda: Decapoda). Mar Biol 136:217–222.

74. Van den Branden C, Richard A, Lemaire J, Decleir W. 1978. La glande nidamentaire accessoire de *Sepia officinalis* L.: analyses biochimiques des pigments des bacteries symbiotiques. Ann Soc r Zool Belg 108:123–139.

75. Lum-Kong A, Hastings TS. 1992. The accessory nidamental glands of *Loligo forbesi* (Cephalopoda: Loliginidae) characterization of symbiotic bacteria and preliminary experiments to investigate factors controlling sexual maturation. J Zool 228:395–403.

76. Kaufman MR, Ikeda Y, Patton C, Van Dykhuizen G, Epel D. 1998. Bacterial Nidamental Symbionts Colonize the Accessory Gland of the Squid *Loligo opalescens* via Horizontal Transmission. Biol Bull 194:36–43.

77. Kerwin AH. 2017. Stability, Development, and Function of a Symbiotic Bacterial Community Associated with the Reproductive System of the Hawaiian Bobtail Squid, *Euprymna scolopes*. University of Connecticut.

78. Fields WG. 1965. The structure, development, food relations, reproduction, and life history of squid *Loligo opalescens* Berry. Calif Dep Fish Game Fish Bull 131:1–108.

79. Smale MJ, Sauer WH, Hanlon RT. 1995. Attempted ambush predation on spawning squids Loligo vulgaris reynaudii by benthic pyjama sharks, Poroderma africanum, of South Africa. J Mar Biol Assoc United Kingdom 7:739–742.

80. Sauer W, Smale M. 1993. Spawning behaviour of *Loligo vulgaris reynaudii* in shallow coastal waters of the South-Eastern Cape, South Africa, p. 489–498. *In* Recent Advances in Cephalopod Fisheries Biology: Contributed Papers to 1991 CIAC International Symposium and Proceedings of the Workshop on Age, Growth.

81. Boletzky SV. 1986. Encapsulation of cephalopod embryos: a search for functional correlations. Am Malacol Bull 4:217–227.

82. Barbieri E, Barry K, Child A, Wainwright N. 1997. Antimicrobial activity in the microbial community of the accessory nidamental gland and egg cases of *Loligo pealei* (Cephalopoda: Loliginidae). Biol Bull 193:275–276.

83. Sherief PM, George MC, Rajasekharan J, Nair DP, Joseph SM, Senan VP. 2004. Antibacterial activity in the extract of accessory nidamental glands of squid and cuttlefish, p. 47–51. *In* Proceedings of MBR 2004 National Seminar on New Frontiers in Marine Bioscience Research.

84. Gomathi P, Nair P, Sherief PM. 2010. Antibacterial activity in the accessory nidamental gland extracts of the Indian squid, *Loligo duvauceli* Orbigny. Indian J Mar Sci 39:100–104.

85. Gromek SM, Suria A, Fullmer MS, Garcia JL, Gogarten JP, Nyholm S V, Balunas MJ. 2016. *Leisingera* sp. JC1, a bacterial isolate from Hawaiian bobtail squid eggs, produces indigoidine and differentially inhibits vibrios. Front Microbiol 7.

86. Atkinson BG. 1973. Squid nidamental gland extract: isolation of a factor inhibiting ciliary activity. J Exp Zool 184:335–340.

87. McFall-Ngai MJ. 2014. The Importance of Microbes in Animal Development: Lessons from the Squid-Vibrio Symbiosis. Annu Rev Microbiol 68:177–194.

88. Mandel MJ, Dunn AK. 2016. Impact and influence of the natural *Vibrio*-squid symbiosis in understanding bacterial-animal interactions. Front Microbiol 7:1982.

89. Nyholm S V, McFall-Ngai MJ. 2004. The winnowing: establishing the squid-*vibrio* symbiosis. Nat Rev Microbiol 2:632–42.

90. Jones BW, Nishiguchi MK. 2004. Counterillumination in the Hawaiian bobtail squid, *Euprymna scolopes* Berry (Mollusca: Cephalopoda). Mar Biol 144:1151–1155.

91. Lee K, Ruby EG. 1992. Detection of the light organ symbiont, *Vibrio fischeri*, in hawaiian seawater by using lux gene probes. Appl Environ Microbiol 58:942–947.

92. McFall-Ngai MJ, Ruby EG. 1991. Symbiont recognition and subsequent morphogenesis as early events in an animal- bacterial mutualism. Science (80-) 254:1491–1494.

93. Nyholm S V., Stewart JJ, Ruby EG, McFall-Ngai MJ. 2009. Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. Environ Microbiol 11:483–493.

94. McFall-Ngai M, Nyholm S V., Castillo MG. 2010. The role of the immune system in the initiation and persistence of the *Euprymna scolopes*-*Vibrio fischeri* symbiosis. Semin Immunol 22:48–53.

95. Norsworthy AN, Visick KL. 2013. Gimme shelter: how Vibrio fischeri successfully navigates an animal's multiple environments. Front Microbiol 4:1–6.

96. Kremer N, Philipp EER, Carpentier MC, Brennan C a., Kraemer L, Altura M a., Augustin R, Häsler R, Heath-Heckman E a C, Peyer SM, Schwartzman J, Rader B a., Ruby EG, Rosenstiel P, McFall-Ngai MJ. 2013. Initial symbiont contact orchestrates host-organ-wide transcriptional changes that prime tissue colonization. Cell Host Microbe 14:183–194.

97. Aschtgen M-S, Wetzel K, Goldman W, McFall-Ngai M, Ruby E. 2015. *Vibrio fischeri*-derived outer membrane vesicles trigger host development. Cell Microbiol 1–12.

98. Singley CT. 1938. *Euprymna scolopes*, p. 69–74. *In* Cephalopod Life Cycles Vol. 1.

99. Hanlon RT, Claes MF, Ashcraft SE, Dunlap P V. 1997. Laboratory Culture of the Sepiolid Squid *Euprymna scolopes*: A Model System for Bacteria-Animal Symbiosis. Biol Bull 192:364–374.

100. Lee PN, McFall-Ngai MJ, Callaerts P, de Couet HG. 2009. The hawaiian bobtail squid (*Euprymna scolopes*): A model to study the molecular basis of eukaryote-prokaryote

mutualism and the development and evolution of morphological novelties in cephalods. Cold Spring Harb Protoc 4:1–9.

101. Wei SL, Young RE. 1989. Development of symbiotic bacterial bioluminescence in a nearshore cephalopod, *Euprymna scolopes*. Mar Biol 103:541–546.

102. Belcaid M, Casaburi G, Mcanulty SJ, Schmidbaur H, Suria AM, Moriano-Gutierrez S, Pankey MS, Oakley TH, Kremer N, Koch EJ, Collins AJ, Nguyen H, Lek S, Goncharenko-Foster I, Minx P, Sodergren E, Weinstock G, Rokhsar DS, McFall-Ngai M, Simakov O, Foster JS, Nyholm S V. 2019. Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. Proc Natl Acad Sci 116:3030–3035.

103. Collins AJ, Fullmer MS, Gogarten JP, Nyholm S V. 2015. Comparative genomics of Roseobacter clade bacteria isolated from the accessory nidamental gland of *Euprymna scolopes*. Front Microbiol 6:1–14.

104. Wagner M, Horn M. 2006. The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. Curr Opin Biotechnol 17:241–249.

105. Tokuda G, Mikaelyan A, Fukui C, Matsuura Y, Watanabe H, Fujishima M. 2018. Fiber-associated spirochetes are major agents of hemicellulose degradation in the hindgut of wood-feeding higher termites. Proc Natl Acad Sci 115:E11996–E12004.

106. Peterson BF, Scharf ME. 2016. Metatranscriptome analysis reveals bacterial symbiont contributions to lower termite physiology and potential immune functions. BMC Genomics 17:772.

107. Moitinho-silva L, Díez-vives C, Batani G, Esteves AI, Jahn MT, Thomas T. 2017. Integrated metabolism in sponge – microbe symbiosis revealed by genome-centered metatranscriptomics. ISME 11:1651–1666.

108. Weisz JB, Voolstra CR, Fiore CL, Labrie M, Jarett JK, Lesser MP. 2015. Transcriptional activity of the giant barrel sponge, *Xestospongia muta* Holobiont: molecular evidence for metabolic interchange. Front Microbiol 6:1–18.

109. Lee FJ, Rusch DB, Stewart FJ, Mattila HR, Newton ILG. 2015. Saccharide breakdown and fermentation by the honey bee gut microbiome. Environ Microbiol 17:796–815.

110. Hamilton PT, Leong JS, Koop BF, Perlman SJ. 2014. Transcriptional responses in a *Drosophila* defensive symbiosis. Mol Ecol 23:1558–1570.

111. Sobolevskaya MP, Smetanina OF, Speitling M, Shevchenko LS, Dmitrenok PS, Laatsch H, Kuznetsova TA, Ivanova EP, Elyakov GB. 2005. Controlling production of brominated cyclic depsipeptides by *Pseudoalteromonas maricaloris* KMM 636T. Lett Appl Microbiol 40:243–248.

112. Speitling M, Smetanina OF, Kuznetsova TA, Laatsch H. 2007. Bromoalterochromides A and A′, unprecedented chromopeptides from a marine *Pseudoalteromonas maricaloris* strain KMM 636T. J Antibiot (Tokyo) 60:36–42.

113. Kalinovskaya NI, Dmitrenok AS, Kuznetsova TA, Frolova GM, Christen R, Laatsch H,

Alexeeva Y V., Ivanova EP. 2008. "*Pseudoalteromonas januaria*" SUT 11 as the source of rare lipodepsipeptides. Curr Microbiol 56:199–207.

114. Tebben J, Motti C, Tapiolas D, Thomas-Hall P, Harder T. 2014. A coralline algal-associated bacterium, *Pseudoalteromonas* strain J010, yields five new korormicins and a bromopyrrole. Mar Drugs 12:2802–2815.

115. Vynne NG, Mansson M, Nielsen KF, Gram L. 2011. Bioactivity, Chemical Profiling, and 16S rRNA-Based Phylogeny of *Pseudoalteromonas* Strains Collected on a Global Research Cruise. Mar Biotechnol 13:1062–1073.

116. Nguyen DD, Wu C-H, Moree WJ, Lamsa A, Medema MH, Zhao X, Gavilan RG, Aparicio M, Atencio L, Jackson C, Ballesteros J, Sanchez J, Watrous JD, Phelan V V., van de Wiel C, Kersten RD, Mehnaz S, De Mot R, Shank EA, Charusanti P, Nagarajan H, Duggan BM, Moore BS, Bandeira N, Palsson BO, Pogliano K, Gutierrez M, Dorrestein PC. 2013. MS/MS networking guided analysis of molecule and gene cluster families. Proc Natl Acad Sci 110:E2611–E2620.

117. Ross AC, Gulland LES, Dorrestein PC, Moore BS. 2015. Targeted Capture and Heterologous Expression of the *Pseudoalteromonas* Alterochromide Gene Cluster in *Escherichia coli* Represents a Promising Natural Product Exploratory Platform. ACS Synth Biol 4:414–420.

118. Vijayasarathy S, Prasad P, Fremlin LJ, Ratnayake R, Salim AA, Khalil Z, Capon RJ. 2016. C3 and 2D C3 Marfey's Methods for Amino Acid Analysis in Natural Products. J Nat Prod 79:421–427.

119. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173:697–703.

120. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: An open-source platform for biological-image analysis. Nat Methods 9:676–682.

121. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

122. Tritt A, Eisen JA, Facciotti MT, Darling AE. 2012. An Integrated Pipeline for de Novo Assembly of Microbial Genomes. PLoS One 7.

123. Aziz RK, Bartels D, Best A a, DeJongh M, Disz T, Edwards R a, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek R a, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.

124. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach M a., Muller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0--a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:1–7.

125. Wu YH, Cheng H, Xu L, Jin X Bin, Wang CS, Xu XW. 2017. Physiological and genomic features of a novel violacein-producing bacterium isolated from surface seawater. PLoS One 12:1–14.

126. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

127. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589.

128. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44:W232–W235.

129. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274.

130. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522.

131. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4 : recent updates and new developments. Nucleic Acids Res 1–4.

132. Grimes DJ, Stemmler J, Hada H, May EB, Maneval D, Hetrick FM, Jones RT, Stoskopf M, Colwell RR. 1984. Vibrio species associated with mortality of sharks held in captivity. Microb Ecol 10:271–282.

133. Gauthier MJ. 1977. *Alteromonas citrea*, a New Gram-Negative, Yellow-Pigmented Species from Seawater. Int J Syst Bacteriol 27:349–354.

134. Park YD, Bak KS, Yi H, Bae KS, Chun J. 2005. *Pseudoalteromonas byusanensis* sp. nov., isolated from tidal flat sediment in Korea. Int J Syst Evol Microbiol 55:2519–2523.

135. Isnansetyo A, Kamei Y. 2003. *Pseudoalteromonas phenolica* sp. nov., a novel marine bacterium that produces phenolic anti-methicillin-resistant *Staphylococcus aureus* substances. Int J Syst Evol Microbiol 53:583–588.

136. Neu AK, Månsson M, Gram L, Prol-García MJ. 2014. Toxicity of bioactive and probiotic marine bacteria and their secondary metabolites in *Artemia* sp. and *Caenorhabditis elegans* as eukaryotic model organisms. Appl Environ Microbiol 80:146–153.

137. Gauthier G, Gauthier M, Christen R. 1995. Phylogenetic analysis of the genera Alteromonas, Shewanella, and Moritella using genes coding for small-subunit rRNA sequences and division of the genus Alteromonas into two genera, Alteromonas (emended) and Pseudoalteromonas gen. nov., and proposal of tw. Int J Syst Bacteriol 45:755–761.

138. Asahina A, Hadfield M. 2015. Draft Genome Sequence of *Pseudoalteromonas luteoviolacea* HI1, Determined Using Roche 454 and PacBio Single-Molecule Real-Time Hybrid Sequencing. Genome Announc 3:4–5.

139. Yu M, Wang J, Tang K, Shi X, Wang S, Zhu WM, Zhang XH. 2012. Purification and characterization of antibacterial compounds of *Pseudoalteromonas flavipulchra* JG1. Microbiology 158:835–842.

140. Sawabe T, Tanaka R, Iqbal MM, Tajima K, Ezura Y, Ivanova EP, Christen R. 2000. Assignment of *Alteromonas elyakovii* KMM 162T and five strains isolated from spot-wounded fronds of *Laminaria japonica* to *Pseudoalteromonas elyakovii* comb. nov. and the extended description of the species. Int J Syst Evol Microbiol 50:265–271.

141. Machado H, Sonnenschein EC, Melchiorsen J, Gram L. 2015. Genome mining reveals unlocked bioactive potential of marine Gram-negative bacteria. BMC Genomics 16:1–12.

142. Xie B Bin, Shu YL, Qin QL, Rong JC, Zhang XY, Chen XL, Shi M, He HL, Zhou BC, Zhang YZ. 2012. Genome sequences of type strains of seven species of the marine bacterium *Pseudoalteromonas*. J Bacteriol 194:2746–2747.

143. Holmstrom C, James S, Neilan BA, White DC. 1998. *Pseudoalteromonas tunicata* sp. nov., a bacterium that produces antifouling agents. Int J Syst Bacteriol 48:1205–1212.

144. Egan S, Holmström C, Kjelleberg S. 2001. *Pseudoalteromonas ulvae* sp. nov., a bacterium with antifouling activities isolated from the surface of a marine alga. Int J Syst Evol Microbiol 51:1499–1504.

145. Xu X-W, Wu Y-H, Wang C-S, Gao X-H, Wang X-G, Wu M. 2010. *Pseudoalteromonas lipolytica* sp. nov., isolated from the Yangtze River estuary. Int J Syst Evol Microbiol 60:2176–81.

146. Matsuyama H, Minami H, Kasahara H, Kato Y, Murayama M, Yumoto I. 2013. *Pseudoalteromonas arabiensis* sp. nov., a marine polysaccharide-producing bacterium. Int J Syst Evol Microbiol 63:1805–1809.

147. Yan J, Wu Y-H, Meng F-X, Wang C-S, Xiong S-L, Zhang X-Y, Zhang Y-Z, Xu X-W, Zhang D-M. 2016. *Pseudoalteromonas gelatinilytica* sp. nov., isolated from surface seawater. Int J Syst Evol Microbiol 66:3538–3545.

148. Matsuyama H, Minami H, Kasahara H, Kato Y, Murayama M, Yumoto I. 2014. *Pseudoalteromonas shioyasakiensis* sp. nov., a marine polysaccharide-producing bacterium. Int J Syst Evol Microbiol 63:1805–1809.

149. Bowman JP. 1998. Pseudoalteromonas prydzensis sp. nov., a psychotrophic, halotolerant bacterium from Antarctic sea ice. Int J Syst Bacteriol 48:1037–1041.

150. Holmström C, Egan S, Franks A, McCloy S, Kjelleberg S. 2002. Antifouling activities expressed by marine surface associated *Pseudoalteromonas* species. FEMS Microbiol Ecol 41:47–58.

151. Romanenko LA, Zhukova N V., Lysenko AM, Mikhailov V V., Stackebrandt E. 2003. Assignment of "*Alteromonas marinoglutinosa*" NCIMB 1770 to *Pseudoalteromonas mariniglutinosa* sp. nov., nom. rev., comb. nov. Int J Syst Evol Microbiol 53:1105–1109.

152. Ivanova EP, Sawabe T, Alexeeva Y V., Lysenko AM, Gorshkova NM, Hayashi K, Zukova N V., Christen R, Mikhailov V V. 2002. *Pseudoalteromonas issachenkonii* sp.

nov., a bacterium that degrades the thallus of the brown alga *Fucus evanescens*. Int J Syst Evol Microbiol 52:229–234.

153. Romanenko LA, Zhukova N V., Rohde M, Lysenko AM, Mikhailov V V., Stackebrandt E. 2003. *Pseudoalteromonas agarivorans* sp. nov., a novel marine agarolytic bacterium. Int J Syst Evol Microbiol 53:125–131.

154. Nam Y Do, Chang HW, Park JR, Kwon HY, Quan ZX, Park YH, Lee JS, Yoon JH, Bae JW. 2007. *Pseudoalteromonas marina* sp. nov., a marine bacterium isolated from tidal flats of the Yellow Sea, and reclassification of *Pseudoalteromonas sagamiensis* as *Algicola sagamiensis* comb. nov. Int J Syst Evol Microbiol 57:12–18.

155. Ivanova EP, Sawabe T, Lysenko AM, Gorshkova NM, Hayashi K, Zhukova N V, Nicolau D V, Christen R, Mikhailov V V, Atcc T. 2002. *Pseudoalteromonas translucida* sp. nov. and *Pseudoalteromonas paragorgicola* sp. nov., and emended description of the genus. Int J Syst Evol Microbiol 52:1759–1766.

156. Bernbom N, Ng YY, Kjelleberg S, Harder T, Gram L. 2011. Marine bacteria from Danish coastal waters show antifouling activity against the marine fouling bacterium *Pseudoalteromonas* sp. strain S91 and zoospores of the green alga *Ulva australis* independent of bacteriocidal activity. Appl Environ Microbiol 77:8557–67.

157. Bowman JP. 2007. Bioactive compound synthetic capacity and ecological significance of marine bacterial genus *Pseudoalteromonas*. Mar Drugs 5:220–41.

158. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Düsterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJN, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kötter P, Krug D, Masschelein J, Melnik A V, Mantovani SM, Monroe EA, Moore M, Moss N, Nützmann H-W, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, Yim G, Yu F, Xie Y, Aigle B, Apel AK, Balibar CJ, Balskus EP, Barona-Gómez F, Bechthold A, Bode HB, Borriss R, Brady SF, Brakhage AA, Caffrey P, Cheng Y-Q, Clardy J, Cox RJ, De Mot R, Donadio S, Donia MS, van der Donk WA, Dorrestein PC, Doyle S, Driessen AJM, Ehling-Schulz M, Entian K-D, Fischbach MA, Gerwick L, Gerwick WH, Gross H, Gust B, Hertweck C, Höfte M, Jensen SE, Ju J, Katz L, Kaysser L, Klassen JL, Keller NP, Kormanec J, Kuipers OP, Kuzuyama T, Kyrpides NC, Kwon H-J, Lautru S, Lavigne R, Lee CY, Linquan B, Liu X, Liu W, Luzhetskyy A, Mahmud T, Mast Y, Méndez C, Metsä-Ketelä M, Micklefield J, Mitchell DA, Moore BS, Moreira LM, Müller R, Neilan BA, Nett M, Nielsen J, O'Gara F, Oikawa H, Osbourn A, Osburne MS, Ostash B, Payne SM, Pernodet J-L, Petricek M, Piel J, Ploux O, Raaijmakers JM, Salas JA, Schmitt EK, Scott B, Seipke RF, Shen B, Sherman DH, Sivonen K, Smanski MJ, Sosio M, Stegmann E, Süssmuth RD, Tahlan K, Thomas CM, Tang Y, Truman AW, Viaud M, Walton JD, Walsh CT, Weber T, van Wezel GP, Wilkinson B, Willey JM, Wohlleben W, Wright GD, Ziemert N, Zhang C, Zotchev SB, Breitling R, Takano E, Glöckner FO. 2015. The Minimum Information about a Biosynthetic Gene cluster (MIBiG) specification. Nat Chem Biol 11:625–631.

159. Ross AC, Gulland LES, Dorrestein PC, Moore BS. 2015. Targeted capture and heterologous expression of the *Pseudoalteromonas alterochromide* gene cluster in *Escherichia coli* represents a promising natural product exploratory platform. ACS Synth Biol 4:414–20.

160. Fountoulakis M, Lahm H-W. 1998. Hydrolysis and amino acid composition analysis of proteins. J Chromatogr A 826:109–134.

161. Offret C, Desriac F, Le Chevalier P, Mounier J, Jégou C, Fleury Y. 2016. Spotlight on antimicrobial metabolites from the marine bacteria Pseudoalteromonas: Chemodiversity and ecological significance. Mar Drugs 14:129.

162. Mikhailov V, Romanenko L, Ivanova E. 2006. The genus *Alteromonas* and related *Proteobacteria*. The Prokaryotes 597–645.

163. Richards GP, Watson MA, Needleman DS, Uknalis J, Boyd EF, Fay P. 2017. Mechanisms for *Pseudoalteromonas piscicida*-Induced Killing of Vibrios and Other Bacterial Pathogens. Appl Environ Microbiol 83:1–17.

164. Miao L, Kwong TFN, Qian PY. 2006. Effect of culture conditions on mycelial growth, antibacterial activity, and metabolite profiles of the marine-derived fungus *Arthrinium* c.f. *saccharicola*. Appl Microbiol Biotechnol 72:1063–1073.

165. Paulsen SS, Andersen B, Gram L, MacHado H. 2016. Biological potential of chitinolytic marine bacteria. Mar Drugs 14:1–17.

166. Skerratt JH, Bowman JP, Hallegraeff G, James S, Nichols PD. 2002. Algicidal bacteria associated with blooms of a toxic dinoflagellate in a temperate Australian estuary. Mar Ecol Prog Ser 244:1–15.

167. Huang S, Hadfield M. 2003. Composition and density of bacterial biofilms determine larval settlement of the polychaete *Hydroides elegans*. Mar Ecol Prog Ser 260:161–172.

168. Tebben J, Tapiolas DM, Motti CA, Abrego D, Negri AP, Blackall LL, Steinberg PD, Harder T. 2011. Induction of larval metamorphosis of the coral *Acropora millepora* by tetrabromopyrrole isolated from a *Pseudoalteromonas* bacterium. PLoS One 6:1–8.

169. Huggett MJ, Williamson JE, Nys R De, Sta V, Peter K. 2006. Larval settlement of the common Australian sea urchin *Heliocidaris erythrogramma* in response to bacteria from the surface of coralline algae. Oecologia 149:604–619.

170. Huang Y, Callahan S, Hadfield MG. 2012. Recruitment in the sea: bacterial genes required for inducing larval settlement in a polychaete worm. Sci Rep 2:228.

171. Jin G, Wang S, Yu M, Yan S, Zhang XH. 2010. Identification of a marine antagonistic strain JG1 and establishment of a polymerase chain reaction detection technique based on the gyrB gene. Aquac Res 41:1867–1874.

172. Yu M, Tang K, Liu J, Shi X, Gulder TA, Zhang X-H. 2013. Genome analysis of *Pseudoalteromonas flavipulchra* JG1 reveals various survival advantages in marine environment. BMC Genomics 14:707.

173. Schleicher TR, Nyholm S V. 2011. Characterizing the host and symbiont proteomes in the association between the bobtail squid, *Euprymna scolopes*, and the bacterium, *Vibrio fischeri*. PLoS One 6:e25649.

174. Nyholm S V., McFall-Ngai M. 2004. The winnowing: establishing the squid–vibrio symbiosis. Nat Rev Microbiol 2:632–642.

175. Visick KL, Ruby EG. 2006. *Vibrio fischeri* and its host: it takes two to tango. Curr Opin Microbiol 9:632–638.

176. Nyholm S V., Mcfall-Ngai MJ. 1998. Sampling the light-organ microenvironment of *Euprymna scolopes*: Description of a population of host cells in association with the bacterial symbiont *Vibrio fischeri*. Biol Bull 195:89–97.

177. Boettcher KJ, Ruby EG, McFall-Ngai MJ. 1996. Bioluminescence in the symbiotic squid Euprymna scolopes is controlled by a daily biological rhythm. J Comp Physiol A 179:65–73.

178. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, Liachko I, Snelling TJ, Dewhurst RJ, Walker AW, Roehe R, Watson M. 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nat Commun 9:10.1038/s41467-018-03317–6.

179. Slaby BM, Hackl T, Horn H, Bayer K, Hentschel U. 2017. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. ISME J 1–14.

180. Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci 109:11002–11007.

181. Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach M a., Weber T, Takano E, Breitling R. 2011. AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:339–346.

182. Li D, Liu C, Luo R, Sadakane K, Lam T. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676.

183. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.

184. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

185. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Data GP, Sam T. 2009. The Sequence Alignment / Map format and SAMtools. Bioinformatics 25:2078–2079.

186. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46.

187. Eren AM, Esen C, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o : an advanced analysis and visualization platform for 'omics data. PeerJ 1–29.

188. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146.

189. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731.

190. Campbell JH, Donoghue PO, Campbell AG, Schwientek P, Sczyrba A. 2013. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc Natl Acad Sci 110:5540–5545.

191. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee Y, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 1–7.

192. Richter M, Rossello-Mora R, Glockner FO, Peplies J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929–931.

193. Spring S, Bunk B, Sproer C, Schumann P, Rohde M, Tindall BJ, Klenk H-P. 2016. Characterization of the first cultured representative of *Verrucomicrobia* subdivision 5 indicates the proposal of a novel phylum. ISME J 10:2801–2816.

194. He S, Stevens S. LR, Chan L-K, Bertilsson S, Glavina del Rio T, Tringe SG, Malmstrom RR, McMahon KD. 2017. Ecophysiology of Freshwater *Verrucomicrobia* Inferred from Metagenome-Assembled Genomes. mSphere 2:1–17.

195. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60.

196. Collins AJ, Nyholm S V. 2011. Draft genome of *Phaeobacter gallaeciensis* ANG1, a dominant member of the accessory nidamental gland of *Euprymna scolopes*. J Bacteriol 193:3397–3398.

197. Blaauwen T Den, Luirink J. 2019. Checks and Balances in Bacterial Cell Division. MBio 10:e00149-19.

198. Liu R, Ochman H. 2007. Stepwise formation of the bacterial flagellar system. Proc Natl Acad Sci 104:7116–7121.

199. Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA. 2012. Control of mucin-type O-

glycosylation: A classification of the polypeptide GalNAc-transferase gene family. Glycobiology 22:736–756.

200. Tailford LE, Crost EH, Kavanaugh D, Juge N. 2015. Mucin glycan foraging in the human gut microbiome. Front Genet 6:doi:10.3389/fgene.2015.00081.

201. Katayama T, Fujita K, Yamamoto K. 2005. Novel Bifidobacterial Glycosidases Acting on Sugar Chains of Mucin Glycoproteins. J Biosci Bioeng 99:457–465.

202. Derrien M, Vaughan EE, Plugge CM, de Vos WM. 2004. Akkermansia municiphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol 54:1469–1476.

203. Berry D, Stecher B, Schintlmeister A, Reichert J, Brugiroux S, Wild B. 2013. Host-compound foraging by intestinal microbiota revealed by single-cell stable isotope probing. Proc Natl Acad Sci 110:4720–4725.

204. Vosseberg J, Martijn J, Ettema JG. 2018. Draft Genome Sequence of "*Candidatus* Moanabacter tarae," Representing a Novel Marine Verrucomicrobial Lineage. Microbiol Announc 7:e00951-18.

205. Zhou L-Y, Wang N-N, Mu D-S, Liu Y, Du Z-J. 2019. *Coraliomargarita sinensis* sp. nov., isolated from a marine solar saltern. Int J Evol Microbiol 69:701–707.

206. Freitas S, Hatosy S, Fuhrman JA, Huse SM, Mark Welch DB, Sogin ML, Martiny AC. 2012. Global distribution and diversity of marine *Verrucomicrobia*. ISME J 6:1499–1505.

207. Hedlund BP, Gosink JJ, Staley JT. 1997. *Verrucomicrobia* div. nov., a new division of the Bacteria containing three new species of *Prosthecobacter*. Antonie Van Leeuwenhoek 72:29–38.

208. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048.

209. Hugenholtz P, Hugenholtz P, Goebel BM, Goebel BM, Pace NR, Pace NR. 1998. Impact of culture independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol v:180p4765-4774.

210. Vandekerckhove TTM, Willems A, Gillis M, Coomans A. 2000. Occurrence of novel verrucomicrobial species, endosymbiotic and associated with parthenogenesis in *Xiphinema americanum*-group species (Nematoda, Longidoridae). Int J Syst Evol Microbiol 50:2197–2205.

211. Schlesner H, Jenkins C, Staley JT. 2006. The Phylum Verrucomicrobia : A Phylogenetically Heterogeneous Bacterial Group, p. 881–896. *In* Prokaryotes.

212. Choo YJ, Lee K, Song J, Cho JC. 2007. *Puniceicoccus vermicola* gen. nov., sp. nov., a novel marine bacterium, and description of *Puniceicoccaceae* fam. nov., *Puniceicoccales* ord. nov., *Opitutaceae* fam. nov., *Opitutales* ord. nov. and *Opitutae* classis nov. Int J Syst Evol Microbiol 57:532–537.

213. Cabello-Yeves PJ, Ghai R, Mehrshad M, Picazo A, Camacho A, Rodriguez-Valera F. 2017. Reconstruction of diverse verrucomicrobial genomes from metagenome datasets of freshwater reservoirs. Front Microbiol 8.

214. Cimermancic P, Medema MH, Claesen J, Kurita K, Brown LCW, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Linington RG, Fischbach MA. 2014. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. Cell 158:412–421.

215. Goldstein S, Beka L, Graf J, Klassen JL. 2019. Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. BMC Genomics 20:doi:10.1186/s12864-018-5381-7.

216. Edlund A, Yang Y, Yooseph S, He X, Shi W, Mclean JS. 2018. Uncovering complex microbiome activities via metatranscriptomics during 24 hours of oral biofilm assembly and maturation. Microbiome 6:1–22.

217. Westermann AJ, Barquist L, Vogel J. 2017. Resolving host–pathogen interactions by dual RNA-seq. PLOS Pathog 1–19.

218. Grote A, Voronin D, Ding T, Twaddle A, Unnasch TR, Lustigman S, Ghedin E. 2017. Defining *Brugia malayi* and *Wolbachia* symbiosis by stage-specific dual RNA-seq. PLoS Negl Trop Dis 11:e0005357.

219. Kopylova E, Noe L, Touzet H. 2012. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217.

220. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–52.

221. Bryant DM, Johnson K, Ditommaso T, Regev A, Haas BJ, Whited JL, Bryant DM, Johnson K, Ditommaso T, Tickle T, Couger MB, Guzikowski AR, Tsai SL, Coyne S, Ye WW, Freeman RM, Peshkin L. 2017. A Tissue-Mapped Axolotl De Novo Transcriptome enables identification of limb regeneration factors. Cell Rep 18:762–776.

222. Louyakis AS, Gourle H, Casaburi G, Bonjawo RME, Duscher AA, Foster JS. 2017. A year in the life of a thrombolite : comparative metatranscriptomics reveals dynamic metabolic changes over diel and seasonal cycles. Environ Microbiol 00:00–00.

223. Huson DH, Beier S, Flade I, Górska A, El-hadidi M. 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data 1–12.

224. Nyholm S V., Deplancke B, Gaskins HR, Apicella M a., McFall-Ngai MJ. 2002. Roles of *Vibrio fischeri* and nonsymbiotic bacteria in the dynamics of mucus secretion during symbiont colonization of the *Euprymna scolopes* light organ. Appl Environ Microbiol 68:5113–5122.

225. Rush JS, Rick PD, Waechter CJ. 1997. Polyisoprenyl phosphate specificity of UDP-GIcNAc: undecaprenyl phosphate N-acetylglucosaminyl 1-P transferase from *E. coli*.

Glycobiology 7:315–322.

226. Lee K-H, Ruby EG. 1994. Effect of the squid host on the abundance and distribution of symbiotic *Vibrio fischeri* in nature. Appl Environ Microbiol 60:1565–1571.

227. Login FH, Balmand S, Vallier A, Vincent-Monégat C, Vigneron A, Weiss-Gayet M, Rochat D, Heddi A. 2011. Antimicrobial peptides keep insect endosymbionts under control. Science (80-) 334:362–365.

228. Chin K, Liesack W, Janssen PH. 2001. *Opitutus terrae* gen. nov., sp. nov., to accommodate novel strains of the division '*Verrucomicrobia*' isolated from rice paddy soil. Int J Syst Evol Microbiol 51:1965–1968.

229. Yoon J, Yasumoto-hirose M, Matsuo Y, Nozawa M, Matsuda S, Kasai H, Yokota A. 2007. *Pelagicoccus mobilis* gen. nov., sp. nov., *Pelagicoccus albus* sp. nov. and *Pelagicoccus litoralis* sp. nov., three novel members of subdivision 4 within the phylum '*Verrucomicrobia*', isolated from seawater by *in situ* culti. Int J Syst Evol Microbiol 57:1377–1385.

230. Hong PY, Croix JA, Greenberg E, Gaskins HR, Mackie RI. 2011. Pyrosequencing-based analysis of the mucosal microbiota in healthy individuals reveals ubiquitous bacterial groups and micro-heterogeneity. PLoS One 6.

231. Sugano Y, Kodama H, Terada I, Yamazaki Y, Noma M. 1994. Purification and characterization of a novel enzyme, α-neoagarooligosaccharide hydrolase (α-NAOS hydrolase), from a marine bacterium, *Vibrio* sp. strain JT0107. J Bacteriol 176:6812–6818.

232. Bomar L, Maltz M, Colston S, Graf J. 2011. Directed culturing of microorganisms using metatranscriptomics. MBio 2:1–8.

233. Schwartzman JA, Ruby EG. 2016. A conserved chemical dialog of mutualism: Lessons from squid and vibrio. Microbes Infect 18:1–10.

234. Gil-Turnes MS, Fenical W. 1992. Embryos of Homarus americanus are Protected by Epibiotic Bacteria. Biol Bull 182:105.

235. Fraune S, Anton-Erxleben F, Augustin R, Franzenburg S, Knop M, Schröder K, Willoweit-Ohl D, Bosch TCGC. 2014. Bacteria–bacteria interactions within the microbiota of the ancestral metazoan Hydra contribute to fungal resistance. ISME J 9:1543–1556.

236. Caldera E, Poulsen M, Suen G, Currie C. 2009. Insect symbioses: a case study of past, present, and future fungus-growing ant research. Environ Entomol 38:78–92.

237. Lopanik N. 2014. Chemical defensive symbioses in the marine environment. Funct Ecol 28:328–340.

238. Goecke F, Labes A, Wiese J, Imhoff JF. 2010. Chemical interactions between marine macroalgae and bacteria. Mar Ecol Prog Ser 409:267–300.

239. Kerwin AH, Nyholm S V. 2018. Reproductive system symbiotic bacteria are conserved

between two distinct populations of *Euprymna scolopes* from Oahu, Hawaii. mSphere 3:1–9.

240. Drew G. 1911. Sexual activities of the squid, *Loligo pealii* (Les.) I. Copulation, egg-laying and fertilization. J Morphol 22:327–359.

241. Gomathi P, Nair J, Sherief P. 2010. Antibacterial activity in the accessory nidamental gland extracts of the Indian squid, Loligo duvauceli Orbigny. Indian J Mar Sci 39:100–104.

242. Klassen JL. 2018. Defining microbiome function. Nat Microbiol 3:864–869.

243. O'Donnell K, Sutton DA, Fothergill A, McCarthy D, Rinaldi MG, Brandt ME, Zhang N, Geiser DM. 2008. Molecular phylogenetic diversity, multilocus haplotype nomenclature, and in vitro antifungal resistance within the *Fusarium solani* species complex. J Clin Microbiol 46:2477–2490.

244. Swofford D. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods), Version 4.0 b10 (Sinauer Associates, Sunderland, MA).

245. Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion GARLI 2.01 (available online at https://www.phylo.org).

246. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, p. 1–8. *In* Gateway Computing Environments (GCE) Workshop.

247. Diener AC, Ausubel FM. 2005. Resistance to *Fusarium oxysporum* 1, a dominant *Arabidopsis* disease-resistance gene, is not race specific. Genetics 171:305–321.

248. Benjamino J, Graf J. 2016. Characterization of the core and caste-specific microbiota in the termite, Reticulitermes flavipes. Front Microbiol 7:1–14.

249. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. 2014. Analysis, optimization and verification of illumina-generated 16s rRNA gene amplicon surveys. PLoS One 9.

250. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pẽa AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods.

251. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, Minchin PR, O'hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Szoecs E, Maintainer HW. 2017. Vegan: Community Ecology Package.

252. Sánchez S, Chávez A, Forero A, García-Huante Y, Romero A, Sánchez M, Rocha D, Sánchez B, Avalos M, Guzmán-Trampe S, Rodríguez-Sanoja R, Langley E, Ruiz B. 2010. Carbon source regulation of antibiotic production. J Antibiot (Tokyo) 63:442–459.

253. Espinel-Ingroff A, Kerkering TM. 1991. Spectrophotometric method of inoculum

preparation for the in vitro susceptibility testing of filamentous fungi. J Clin Microbiol 29:393–4.

254. Clinical and Laboratory Standards Institute. 2008. Reference method for broth dilution antifungal susceptibility testing of filamentous fungi (Wayne, PA) 2nd Ed.

255. Chambers MC, MacLean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak MY, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J, Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P. 2012. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol.

256. Wang M, Carver JJ, Phelan V V., Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik A V., Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, Boya CAP, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson B, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol.

257. Shannon P, Lejay L V, Dean A, Chiaromonte F, Shasha DE, Coruzzi GM, Amin N, Schwikowski B, Ideker T, Stitt M. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res.

258. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, Palermo A, Warth B, Hermann G, Koellensperger G, Huan T, Uritboonthai W, Aisporna AE, Wolan DW, Spilker ME, Benton HP, Siuzdak G. 2018. METLIN: A Technology Platform for Identifying Knowns and Unknowns. Anal Chem 90:3156–3164.

259. Short DPG, O'Donnell K, Thrane U, Nielsen KF, Zhang N, Juba JH, Geiser DM. 2013. Phylogenetic relationships among members of the Fusarium solani species complex in human infections and the descriptions of F. keratoplasticum sp. nov. and F. petroliphilum stat. nov. Fungal Genet Biol 53:59–70.

260. Short DPG, O'Donnell K, Geiser DM. 2014. Clonality, recombination, and hybridization in the plumbing-inhabiting human pathogen *Fusarium keratoplasticum* inferred from multilocus sequence typing. BMC Evol Biol 14:91.

261. O'Donnell K, Sutton D a., Rinaldi MG, Sarver B a. J, Balajee S a., Schroers H-J, Summerbell RC, Robert V a. RG, Crous PW, Zhang N, Aoki T, Jung K, Park J, Lee Y-H, Kang S, Park B, Geiser DM. 2010. Internet-Accessible DNA Sequence Database for Identifying Fusaria from Human and Animal Infections. J Clin Microbiol 48:3708–3718.

262. Richards T a., Jones MDM, Leonard G, Bass D. 2012. Marine Fungi: Their Ecology and Molecular Diversity. Ann Rev Mar Sci 4:495–522.

263. Al-Hatmi A, Curfs-Breuker I, de Hoog G, Meis J, Verweij P. 2017. Antifungal Susceptibility Testing of *Fusarium*: A Practical Approach. J Fungi 3:19.

264. Wang M, Carver J, Phelan V, Sanchez L, Garg N, Peng Y, Nguyen D, Watrous J, Kapono C. 2016. Sharing and community curation of mass spectrometry data with GNPS. Nat Biotechnol 34:828–837.

265. Mason D, Dietz A, Deboer C. 1962. Lincomycin, a new antibiotic. I. Discovery and biologic properties. Antimicrob Agents Chemother 555–559.

266. Barber M, Waterworth P. 1964. Antibacterial Activity of Lincomycin and Pristinamycin: A Comparison with Erythromycin. Br Med J 2:603–609.

267. Zhao Q, Wang M, Xu D, Zhang Q, Liu W. 2014. Metabolic coupling of two small-molecule thiols programs the biosynthesis of lincomycin A. Nature 518:115–119.

268. Peschke U, Schmidt H, Zhang H -Z, Piepersberg W. 1995. Molecular characterization of the lincomycin-production gene cluster of *Streptomyces lincolnensis* 78-11. Mol Microbiol 16:1137–1156.

269. Satoi S, Muto N, Hayashi M, Fujii T, Otani M. 1980. Mycinamicins, new macrolide antibiotics. I. Taxonomy, production, isolation, characterization, and properties. J Antibiot 33:364–376.

270. Hayashi M, Kinoshita K, Sudate Y, Satoi S, Sakakibara H. 1983. Mycinamicins, New Macrolide Antibiotics. VII. Structures of Minor Components, Mycinamicin VI and VII. J Antibiot (Tokyo) 36:175–178.

271. Hayashi M, Ohno M, Katsumata S, Satoi S. 1981. Mycinamicins, New Macrolide Antibiotics. IV. Structure of Mycinamicin III. J Antibiot (Tokyo) 34:276–281.

272. Xu H, Valenzuela N, Fai S, Figeys D, Bennett SAL. 2013. Targeted lipidomics - Advances in profiling lysophosphocholine and platelet-activating factor second messengers. FEBS J 280:5652–5667.

273. Galtier d'Auriac I, Quinn RA, Maughan H, Nothias L-F, Little M, Kapono CA, Cobian A, Reyes BT, Green K, Quistad SD, Leray M, Smith JE, Dorrestein PC, Rohwer F, Deheyn DD, Hartmann AC. 2018. Before platelets: the production of platelet-activating factor during growth and stress in a basal marine organism. Proceedings Biol Sci 285.

274. Quinn RA, Vermeij MJA, Hartmann AC, Galtier I, Benler S, Haas A, Quistad SD, Lim YW, Little M, Sandin S, Smith JE, Dorrestein PC, Rohwer F, Quinn RA. 2016. Metabolomics of reef benthic interactions reveals a bioactive lipid involved in coral defence. Proc R Soc London B Biol Sci 283:20160469.

275. Sohlenkamp C, López-Lara IM, Geiger O. 2003. Biosynthesis of phosphatidylcholine in bacteria. Prog Lipid Res 42:115–162.

276. Geiger O, López-Lara IM, Sohlenkamp C. 2013. Phosphatidylcholine biosynthesis and function in bacteria. Biochim Biophys Acta - Mol Cell Biol Lipids 1831:503–513.

277. Williams LW. 1909. The Anatomy of the Common Squid Loligo pealii, LesueurThe American Museum of Natural History. The American Museum of Natural History, New York, NY.

278. Pierantoni U. 1918. Organi luminosi, organi simbiotici e ghiandola nidamentale accessoria nei cephalopodi. Boll della Soc dei Nat Napoli 30:30–36.

279. Wampach L, Heintz-Buschart A, Hogan A, Muller EEL, Narayanasamy S, Laczny CC, Hugerth LW, Bindl L, Bottu J, Andersson AF, de Beaufort C, Wilmes P. 2017. Colonization and Succession within the Human Gut Microbiome by Archaea, Bacteria, and Microeukaryotes during the First Year of Life. Front Microbiol 8:738.

280. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H, Khan MT, Zhang J, Li J, Xiao L, Al-Aama J, Zhang D, Lee YS, Kotowska D, Colding C, Tremaroli V, Yin Y, Bergman S, Xu X, Madsen L, Kristiansen K, Dahlgren J, Wang J, Jun W. 2015. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe 17:690–703.

281. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. 2011. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A 108:4578–85.

282. Jonsbu E, McIntyre M, Nielsen J. 2002. The influence of carbon sources and morphology on nystatin production of *Streptomyces noursei*. J Biotechnol 95:133–144.